

**Model-Based Estimates of Income for Middle Layer
Super Output Areas, 2004/05
Validation Report**

Nicola White

Karen Gask

Shayla Goldring

Jane Longhurst

Marie Cruddas

Samantha Revis

Contents

SUMMARY	1
1 BACKGROUND.....	4
2 VALIDATION AGAINST ALTERNATIVE INCOME DATA SOURCES.....	6
2.1 INTRODUCTION	6
2.2 RESULTS	7
2.2.1 <i>Her Majesties Revenue and Customs data</i>	7
2.2.2 <i>Index of Multiple Deprivation, England</i>	11
2.2.3 <i>Index of Multiple Deprivation, Wales</i>	17
2.2.4 <i>Annual Survey of Hours and Earnings</i>	19
2.2.5 <i>General Household Survey</i>	23
2.3 CONCLUSIONS.....	24
A STATISTICAL RESULTS	26
B GRAPHICAL RESULTS.....	29
C BIBLIOGRAPHY	64

Summary

Consultation with users has indicated that there is a widespread requirement to have better information on income at the small area level. Following the Government's decision not to include an income question in the 2001 Census, the Office for National Statistics (ONS) took the lead in exploring alternative measures for the provision of income data for small areas.

The Small Area Income Estimates Project was established with the aim of producing sets of small area estimates of average household income using the modelling techniques developed by the ONS Methodology Directorate. The methodology involves combining Family Resources Survey data with other information (census, benefits, HM Revenue and Customs data, council tax and house prices) known as covariate¹ data available at the small area level and building a modelled relationship. The model can then be used to produce estimates with confidence intervals of average household income for all MSOAs in England and Wales. For more technical information see White et al (2007).

Estimates and confidence intervals for the values of average ward income (2001/02) for all wards in England and Wales were released as experimental² statistics in November 2005. In keeping with National Statistics geographic and statistical policy, ONS is currently focussing on producing estimates of average income for MSOAs. Super Output Areas (SOAs) are a new geographic hierarchy designed to improve the reporting of small area statistics. A range of areas have been developed that are of consistent size and whose boundaries will not change for electoral or other administrative reasons. These areas are built from groups of Output Areas (OAs) used for the 2001 Census. The SOA layers form a hierarchy based on aggregations of Output Areas (OAs), these add firstly to Lower Layer SOAs (LSOA) then to a larger area named Middle Layer SOAs. MSOAs have a mean population of 7,200 and a

¹ A variable (or variables) that is either known or can be estimated relatively accurately and is then used in the estimation of other variables.

² The term 'experimental' is applied to any set of ONS statistics that do not yet meet the rigorous quality standards of National Statistics and/or may be subject to change due to methodological development.

minimum population of 5,000. They are built from groups of LSOAs and constrained by the 2003 local authority boundaries used for 2001 Census outputs.

Estimates are produced for the following four income types:

- total household weekly income (unequalised);
- net household weekly income (unequalised);
- net household weekly income before housing costs (equalised); and
- net household weekly income after housing costs (equalised).

Note equalised income means that the household income values have been adjusted to take into consideration household size and composition.

The release of a more up-to-date series of model-based average weekly household estimates for MSOAs raises the issue of comparability with the ward 2001/02 estimates. Unfortunately, as set out in Chapter 7 of the Technical Report (White et al (2007)) the 2004/05 estimates are only comparable with the 2001/02 estimates in a limited way.

As with previous estimates, a number of diagnostic checks were used to assess the appropriateness of the models used to produce the MSOA level estimates. The same methodology was employed for the 2004/05 estimates as for previous ward estimates and the results show that in general the models are well specified, the assumptions are sound and the methodology is appropriate.

As well as validating the process of making the estimates it was also necessary to validate the estimates themselves. This is of crucial importance in establishing the plausibility of the model-based estimates. The validation of the model-based estimates consisted of analyses comparing the model-based estimates with other sources of income data.

The first section of the report provides some background to the project and details the approach employed to validate the estimates. Section 2 documents the results of

analyses to compare the MSOA-level model-based estimates with other sources of income data.

In general the results show correlation between the model-based estimates and alternative income data sources are moderately strong. Many of the patterns observed have been explained.

1 Background

Consultation with users, including representatives from central and local Government, the academic and business sectors, about requirements for information from the 2001 Census underlined the need for a question on income. In particular, the work substantiated the widespread and increasing demand for detailed information at a range of geographical levels.

Although the Government recognised this need for information on income, concern about the risks to the conduct of the Census meant that the preferred approach was to assess whether or not requirements could be met by using alternative sources of data. In accordance with proposals set out in the Census White Paper, the Government Statistical Service (GSS) set up a working group to investigate the feasibility of various options including:

- using data on the receipt of benefits from the Department of Work and Pensions (DWP); and
- producing modelled income data.

The results of this work were set out alongside findings from the detailed Census programme of research and question testing in a paper circulated to users so they could identify the preferred approach for meeting their requirements. The Government considered all of this information before they made the decision not to include an income question in the 2001 Census.

Following the initial work into the feasibility of producing model-based estimates of income the ONS established a project to implement this approach. As a result of this project, estimates of the values of average income (2001/02) for wards in England and Wales were released as experimental statistics on the Neighbourhood Statistics website in 2005. The methodology used to produce the estimates is documented in the technical report for this project (Longhurst et al (2005)).

One of the main limitations of the 2001/02 CAS ward estimates highlighted by users was the requirement for estimates of average weekly household income on Super Output Areas (SOAs). Super Output Areas are a new geographic hierarchy designed to improve the reporting of small area statistics. A range of areas have been developed that are of consistent size and whose boundaries will not change for electoral or other administrative reasons. These areas are built from groups of Output Areas (OAs) used for the 2001 Census. The SOA layers form a hierarchy based on aggregations of Output Areas (OAs), these add firstly to Lower Layer SOAs (LSOA) then to a larger area named Middle Layer SOAs (MSOA). MSOA have a mean population of 7,200 and a minimum population of 5,000. They are built from groups of LSOA and constrained by the 2003 local authority boundaries used for 2001 Census outputs. These boundaries offer an advantage over CAS wards because they will not change for electoral or other administrative reasons. This means fairer comparisons can be made between areas and over time, this being another request of users.

A number of diagnostic checks have been used to assess the appropriateness of the models developed for producing MSOA-level estimates of income. The analyses show that in general the models are well specified and the assumptions are sound (Longhurst et al (2005)). This provides confidence in the accuracy of the estimates and the associated confidence intervals. In addition, the methodology and its application to the up-to-date data sources used to produce the 2004/05 model-based estimates has undergone a review by ONS' Methodology Directorate. These processes have ensured that the models developed are the best possible for the data available, assessed the goodness of fit and validated the methodology used to produce the estimates.

As well as validating the process of making the estimates it is necessary to validate the estimates themselves. This is of crucial importance in establishing the plausibility of the model-based estimates. This was carried out by comparing the model-based estimates with other sources of income data, the details of which are contained in this report.

2 Validation Against Alternative Income Data Sources

2.1 Introduction

Model-based MSOA estimates of average weekly household income (four types) for 2004/05 have been produced for all MSOAs in England and Wales. Data for average weekly household income are not available at the MSOA-level; if they were then there would be no need for the production of the model-based estimates. However, alternative income data sources do exist. These data sources may not exactly describe average weekly household income or may not be at the MSOA-level but they all provide indicators of levels of income in an area. They can therefore be used for general comparisons to validate the income estimates and to establish whether the relationship between the model-based estimates and the alternative data sources exhibit the patterns expected across areas. A number of different sources of income data were investigated including:

- HM Revenue and Customs data, 2004/05, England and Wales;
- Index of Multiple Deprivation, 2004, England;
- Index of Multiple Deprivation, 2005, Wales;
- Annual Survey of Hours and Earnings data, 2005, England & Wales;
- General Household Survey data, 2005, England & Wales.

A range of analyses, using these data, were carried out for total income measures. Section 2.2 contains the results of the analyses for total weekly household income.

Analyses between the model-based estimates and the data sources are made graphically comparing actual values and rankings and statistically using Pearson's¹ and Spearman's² correlation coefficients. Comparisons are made across a country and, if appropriate, for a region.

¹ Pearson's correlation coefficient is a measure of linear association. The coefficient takes values between -1 and 1, the further from zero, the stronger the relationship.

² The Spearman Rank correlation coefficient summarises the strength of the association between 2 ranked variables. The coefficient takes values between -1 and +1, the further from zero, the stronger the relationship.

As the alternative sources of income are not available at MSOA-level the model-based MSOA estimates of income are aggregated to the appropriate level i.e. Local Authority District (LAD) level or Government Office Region/Country (GOR) level.

2.2 Results

2.2.1 Her Majesty's Revenue and Customs data

Her Majesties Revenue and Customs Survey of Personal Incomes (SPI) (HMRC (2004/05)) is based on a sample of information held by the HMRC tax offices on persons who could be liable to UK tax. It is carried out annually and covers the income assessable for tax each year. The SPI provides estimates of average yearly total income for UK taxpayers at the Local Authority District (LAD) level. For the Isle of Scilly the data are suppressed due to small sample size. Comparisons were made for 375 out of 376 LADs in England and Wales. The HMRC data were divided by 52 to get a weekly figure comparable with the model-based estimates and the model-based estimates were aggregated to the LAD level.

The model-based estimates describe household income as an average for the total population of an area. The HMRC estimates measure individual income as an average for UK tax payers in an LAD. Both the model-based and HM Revenue and Customs estimates use total income as their definition of income, this is gross income plus any income from benefits in the form of taxes such as Disability Tax Credit and Working Families Tax Credit. Income for individuals is generally less than for households. Here we are comparing average household income with average income for individuals, thus we would generally expect the household income estimates to be larger than the individual income estimates.

Another difference between the two measures of income concerns the denominators used for the calculation of an average value. The model-based estimates are averaged for households and then over all private households in the area. The HMRC estimates are calculated as an average over only UK taxpayers. This implies that in areas of high unemployment or part-time employment, e.g. where there are a high proportion

of people who do not earn enough to pay tax, we may expect to see the average individual income being larger than the household income.

Figure 1 and Figure 2 provide an overall comparison of the two data sources for England and Wales.

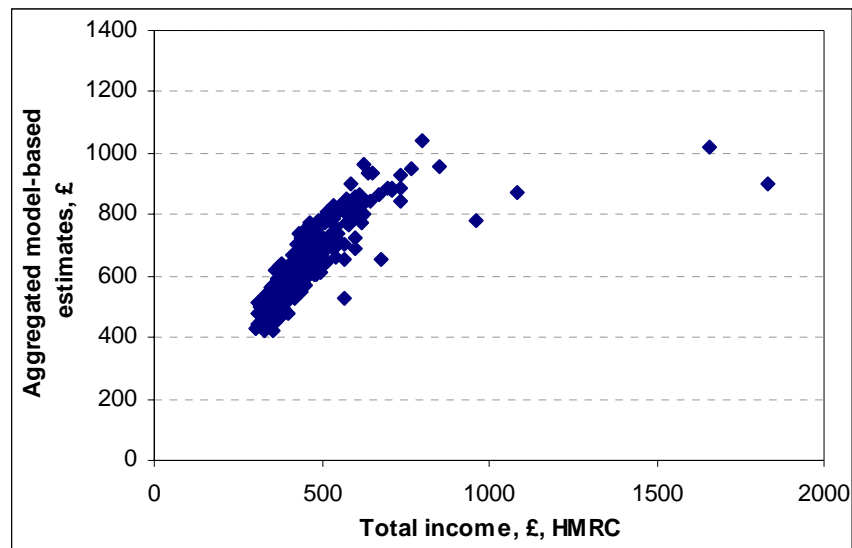


Figure 1: Aggregated model-based estimates compared with HMRC data, England and Wales

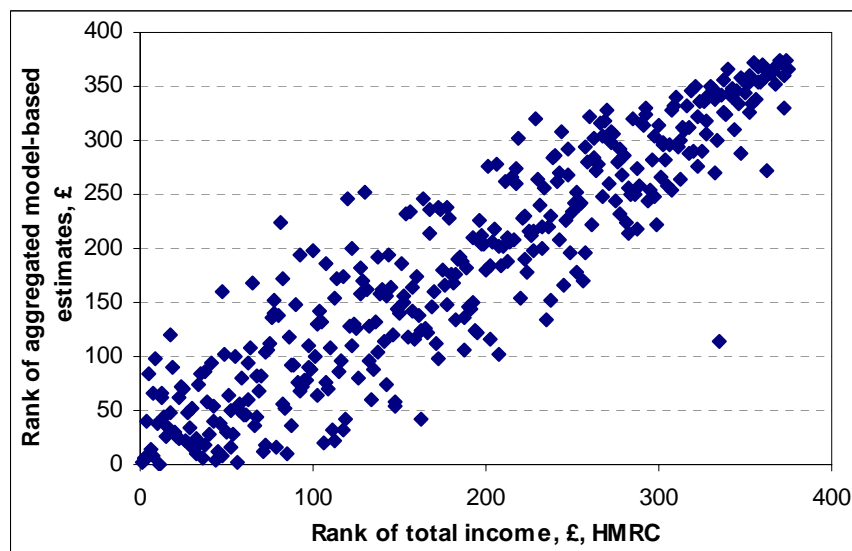


Figure 2: Rank of aggregated model-based estimates compared with rank of HMRC data, England and Wales

Figure 1 shows in general, as expected, the average household income for an LAD is greater than the average individual income in England and Wales. The correlation coefficient for the data is 0.76. Figure 2 compares the ranks of the two data sets and shows a strong positive relationship, summarised by a rank correlation coefficient of 0.92.

The four outliers in Figure 1 represent the values for Kensington and Chelsea, Westminster, Tower Hamlets and City of London LADs in London. These areas are associated with lower average household sizes and contain MSOAs with both low and high average MSOA incomes. Hence individual earnings for UK taxpayers might be expected to be higher than average household income over the total population for these areas.

Figure 3 and Figure 4 display the results for London. Here we see the most variability in results.

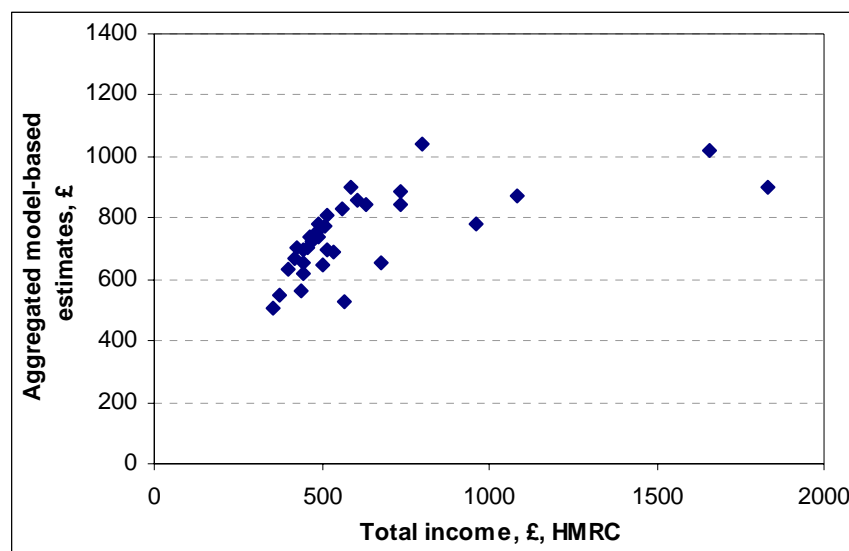


Figure 3: Aggregated model-based estimates compared with HMRC data, London

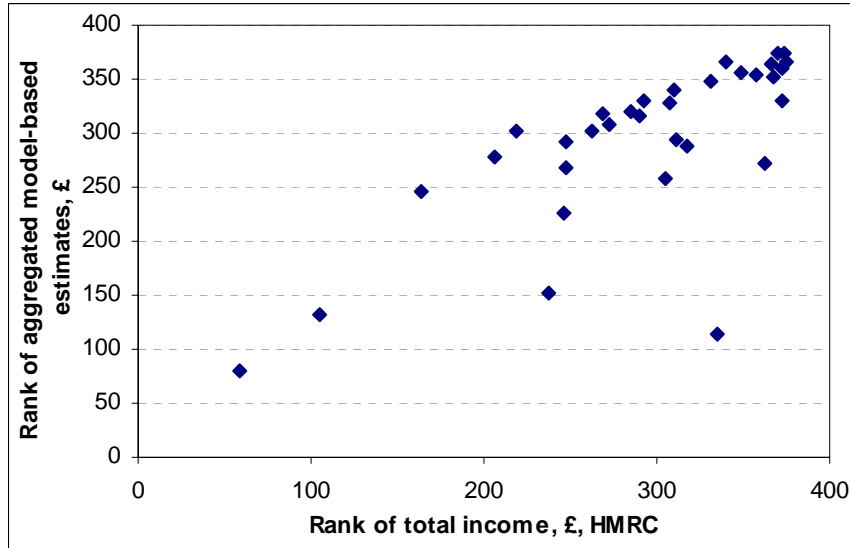


Figure 4: Rank of aggregated model-based estimates compared with rank of HMRC data, London

Figure 3 shows that with the exception of the four outliers discussed earlier, the LADs have similar average household and individual incomes. This is supported by a Pearson correlation coefficient of 0.64. Figure 4 shows LADs ranked highly according to individual income are also ranked highly by average household income. The Spearman Rank correlation coefficient is 0.76, the lowest coefficient for all the regions in England.

The remaining regions exhibit very strong correlations between individual and household income with Pearson correlation coefficients ranging from 0.87 to 0.94. The differences are attributable to differing units of observation employed by the SPI and FRS, i.e. individual vs. household and also differing populations, i.e. UK tax payers vs. private households respectively. Generally, these differences have resulted in lower HMRC estimates.

Appendix A contains a full set of Pearson and Spearman Rank correlations and Appendix B contains a full set of plots for these analyses for total income.

2.2.2 Index of Multiple Deprivation, England

The Index of Multiple Deprivation 2004 (IMD 2004) in England (Department of Communities and Local Government, formerly Office of the Deputy Prime Minister (2004)) is a Lower Layer Super Output Area (LSOA) index, made up of seven domains:

- Income;
- Employment;
- Health deprivation and disability;
- Education, skills and training;
- Barriers to Housing and Services;
- Crime; and
- Living Environment.

Each domain contains a range of indices designed to provide a composite measure of deprivation. The income domain focuses on those people who are living in low income families. The domain score is the proportion of the LSOA population in families in receipt of benefits.

Note a high ranking (1) in the income domain of the IMD 2004 indicates a high proportion of families in receipt of income related benefits.

The following indicators are used for the IMD 2004 income domain:

- Adults and children in households receiving Income Support for 2001. Source: DWP.
- Adults and children in households receiving Income Based Job Seekers Allowance for 2001. Source: DWP.
- Adults and children in households receiving Working Families Tax Credit whose equivalised income (excluding housing benefits) is below 60% of median before housing costs for 2001. Source: HM Revenue and Customs, now Her Majesty's Revenue and Customs.

- Adults and children in Disabled Person's Tax Credit households whose equivalised income (excluding housing benefits) is below 60% of median before housing costs (2001). Source: HM Revenue and Customs, now HMRC.
- National Asylum Support Service supported asylum seekers in England in receipt of subsistence only and accommodation support (2002). Source: Home Office.

The income domain is measured at LSOA level, however, the numbers of income deprived along with population counts and a rank are also provided for LADs. The LAD level summaries have been employed in the analysis for the model-based income estimates in England. The model-based MSOA estimates were combined with estimates of the number of households in MSOAs and aggregated to the LAD level for comparison with the IMD 2004 data.

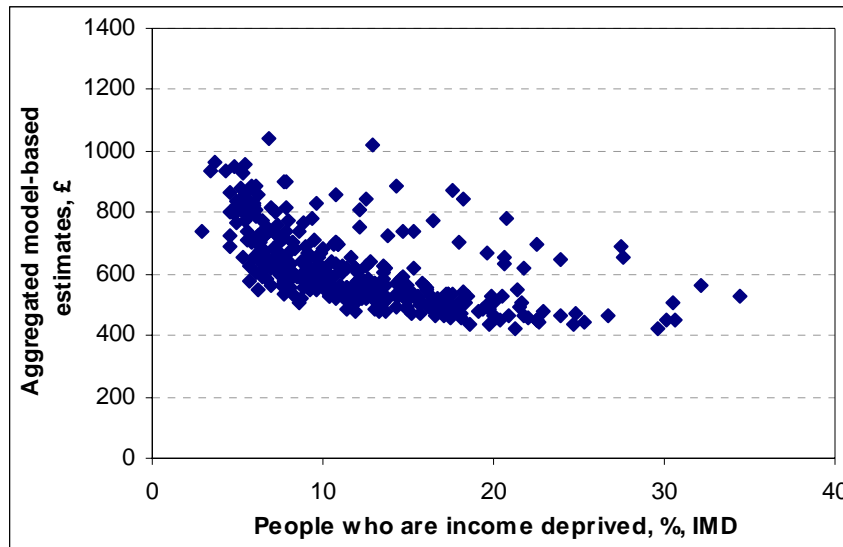


Figure 5: Aggregated model-based MSOA estimates compared with Income Score, IMD, England

The model-based estimates of average total weekly household income were compared with the income score of the IMD 2004 (i.e. the % of families in receipt of various benefits), Figure 5. The Pearson correlation coefficient between the actual values of the two datasets is -0.58 .

The plot shows that a wide range of average income values are associated with low scores on the IMD and that lower levels of average income are spread widely across higher scores on the IMD. The analysis shows that high scores on the IMD can be associated with low values of average income for a LAD. This relationship is possible since the measures are somewhat different. Figure 5 shows that the two different measures do not have a strong linear relationship.

Figure 6 displays a comparison of the rankings of the two income measures. A high ranking (1) for an aggregated model-based estimate indicates a high income. As illustrated in Figure 6 the Spearman Rank correlation coefficient for the data is -0.70 indicating a moderately strong negative relationship. Further analysis by region shows stronger relationships between the IMD income score and the model-based estimates of income.



Figure 6: Rank of Aggregated model-based MSOA estimates compared with rank of Income Score, IMD, England

LADs ranked highly with regard to the size of the model-based estimate (i.e. with a high average LAD income) tend to be ranked low on the Income Score (i.e. a low proportion of people in receipt of income related benefits). One would expect the Spearman Rank correlation coefficient to be negative showing that an LAD ranked

highly on the Income Score would be ranked low on the model-based income scale (an LAD with a lower estimate of average weekly household income).

More information can be obtained by repeating the above analyses by Government Office Region (GOR) in England. Figure 7 and Figure 8 show the correlation between the model-based estimates of income and the Income Score and the rank of the model-based estimates for total weekly household income against the rank of the Income Score for the East Midlands region. The ranking used in Figure 8 is the same as that for England, e.g. an LAD ranked 136th for the whole of England is still ranked 136th in the East Midlands plot. This method of ranking has been preserved to illustrate both the strength of association between the data sources and also the relative wealth/income score of the region in comparison to the rest of England.

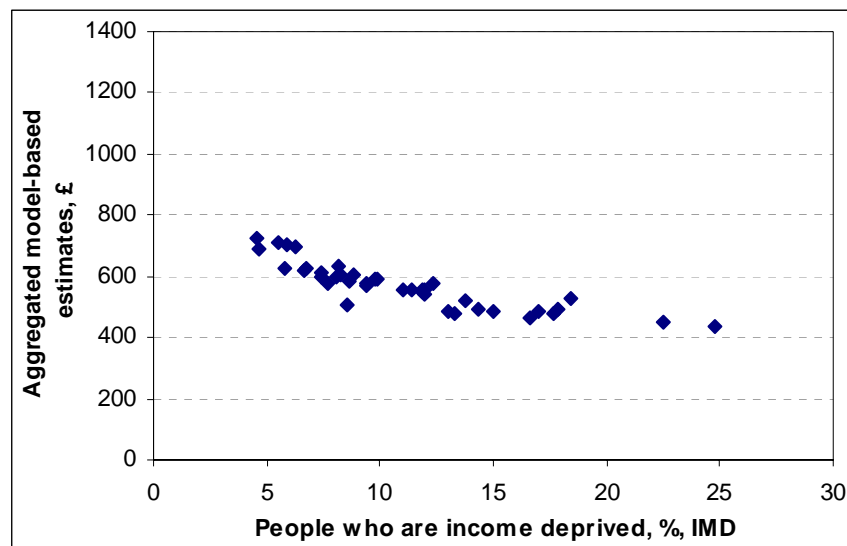


Figure 7: Model-based MSOA estimates compared with Income Score, IMD, East Midlands

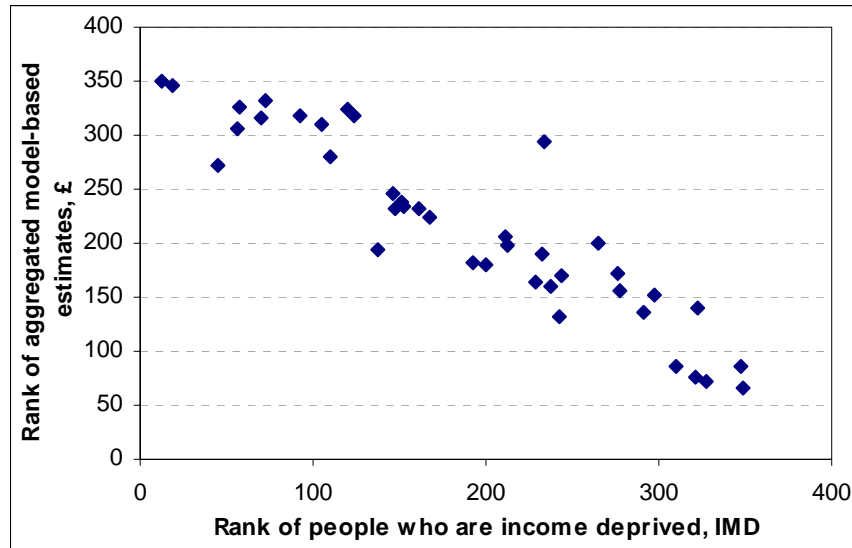


Figure 8: Rank of model-based MSOA estimates compared with rank of Income Score, IMD, East Midlands

The results are as expected. The ranking plot shows a strong negative correlation and the trend in Figure 9 is similar to that seen in Figure 5, but with a slightly stronger negative relationship between average weekly household income and the income score.

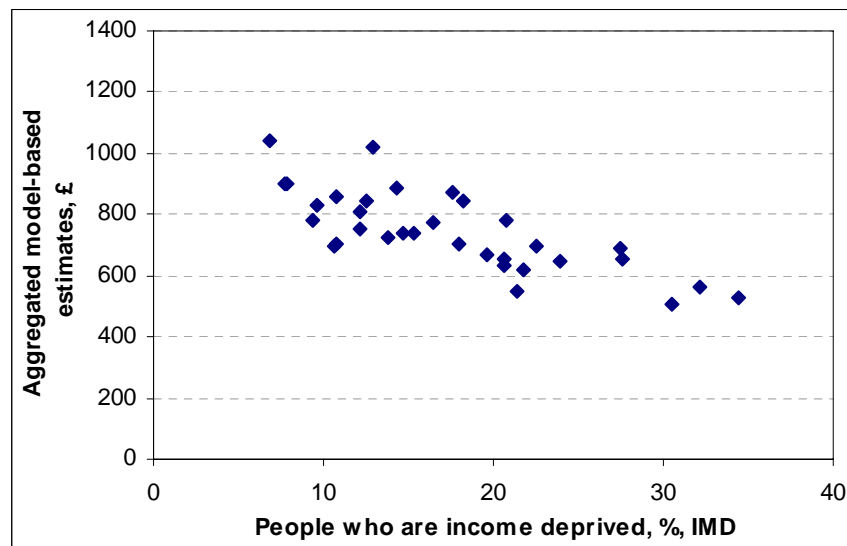


Figure 9: Aggregated model-based MSOA estimates compared with Income Score, IMD, London

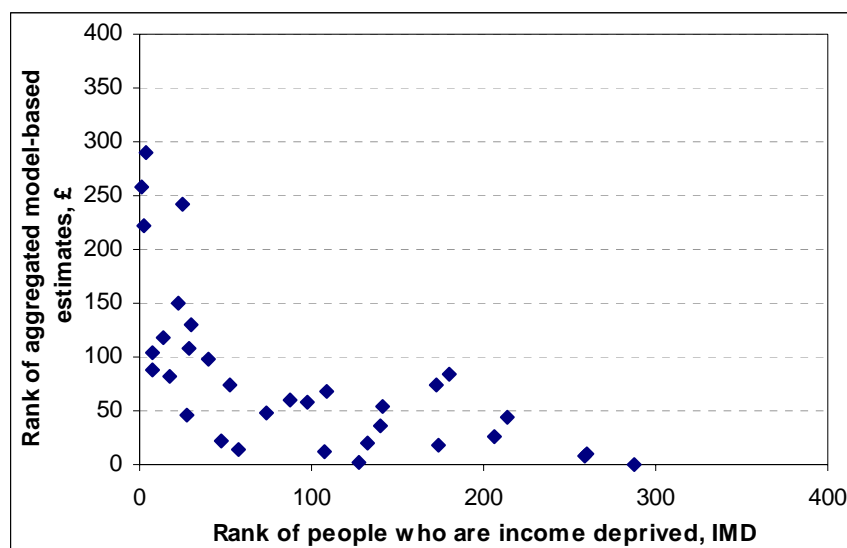


Figure 10: Rank of aggregated model-based MSOA estimates compared with rank of Income Score, IMD, London

Figure 10 shows a negative relationship between the two rankings. LADs that have a low ranking for average weekly household income (as determined by the aggregated model-based estimates across the whole of England) have a high ranking determined by the income score and vice versa. However, Figure 10 does show that the relationship is not always strong in individual cases (for example some LADs are similarly ranked in each dataset).

This is an expected trend for London. In LADs and even within MSOAs there are likely to be areas with very different levels of income. For example, areas of poverty may be very close to areas of wealth, e.g. a MSOA with a high proportion of the population claiming benefits but also a high average income. Therefore, in comparison with other areas of England, the LADs in London will for a particular income score generally be associated with higher levels of income.

Appendix B contains a full set of plots for this analysis for total income and shows that for the majority of the regions the patterns exhibited were the same as those shown for the East Midlands. Some of the ranking plots (West Midlands, North East and Yorkshire and The Humber) identify some outliers where LADs have a low ranking according to their average weekly household income and a low ranking

according to their Income Score. These LADs have low values for average income and low levels of the population claiming benefits.

Appendix A contains a full set of Pearson and Spearman Rank correlation coefficients for this analysis of total income.

Note similar results as those described above for total income were obtained for the three other income types.

2.2.3 Index of Multiple Deprivation, Wales

A separate LSOA-level IMD has been constructed for Wales, (Welsh Assembly (2005)). The Welsh IMD 2005 was produced based on LSOAs, however, the summaries for LADs have been constructed using the estimate of the number of income deprived people from the IMD and the estimate of the number of people 'at risk' of income deprivation from the IMD in each LSOA. These counts were aggregated to LAD level in order to form a percentage of those at risk of income deprivation for each LAD. For the purposes of this analysis, the model-based MSOA estimates for 2004/05 have been aggregated to the LAD level combined with the estimates of the number of households in MSOAs. The following indicators were used for the Welsh IMD Income Domain:

- Adults in households receiving Income Support for 2004.
- Children in households receiving Income Support for 2004.
- Adults in households receiving Income Based Job Seekers Allowance for 2004.
- Children in households receiving Income Based Job Seekers Allowance for 2004.
- Adults in households receiving Working Families Credit below a low income threshold for 2002.
- Children in households receiving Working Families Credit below a low income threshold for 2002.

- Adults in households receiving Disability Tax Credit below a low income threshold for 2002.
- Children in households receiving Disability Tax Credit below a low income threshold for 2002.
- National Asylum Support Service supported asylum seekers in Wales in receipt of subsistence only and accommodation support for 2004.

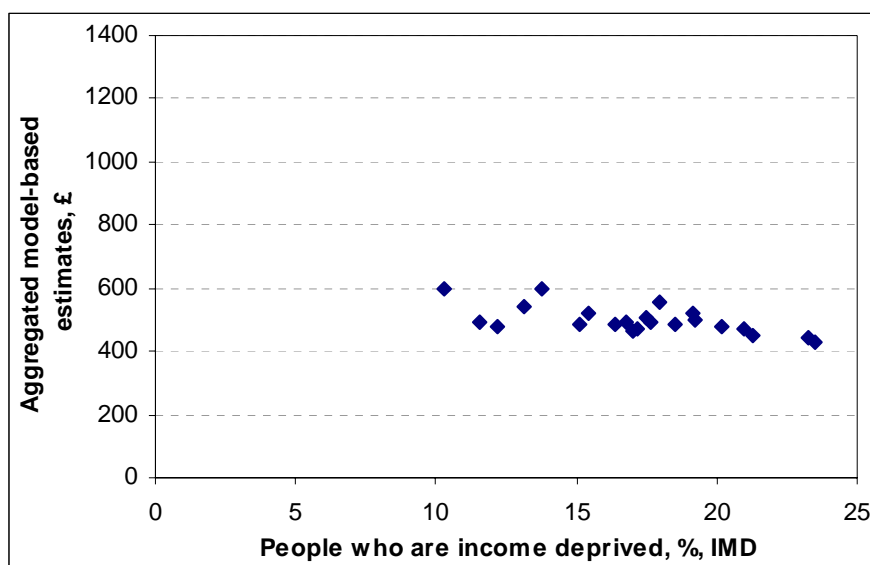


Figure 11: Model-based MSOA estimates compared with Income IMD, Wales

The patterns shown by Figure 11 and Figure 12 for Wales are similar to the results found in England. A negative correlation exists between the rankings as represented by a Spearman Rank correlation coefficient of -0.55.

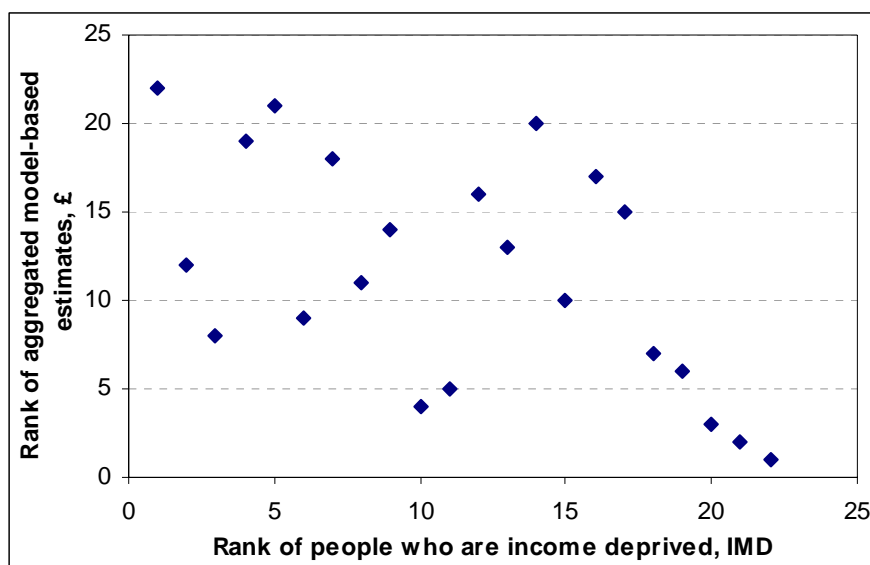


Figure 12: Rank of model-based MSOA estimates compared with rank of Income
IMD, Wales

Note similar results were obtained for the three other income types.

2.2.4 Annual Survey of Hours and Earnings

The Annual Survey of Hours and Earnings (ASHE) for 2005, covering England, Wales, Scotland and Northern Ireland (ONS (2006)) was also considered an appropriate data source to compare against the model-based estimates of average weekly household income. The ASHE is a new survey developed to replace the New Earnings Survey from 2004 and includes data collected from employees. It provides estimates of average gross weekly earnings for adult employees at the LAD level (this does not include self-employment data). The model-based estimates were aggregated to the LAD level and comparisons made.

The model-based estimates describe household income as an average for all private households in an area. The ASHE estimates measure the average earnings for employees in a LAD.

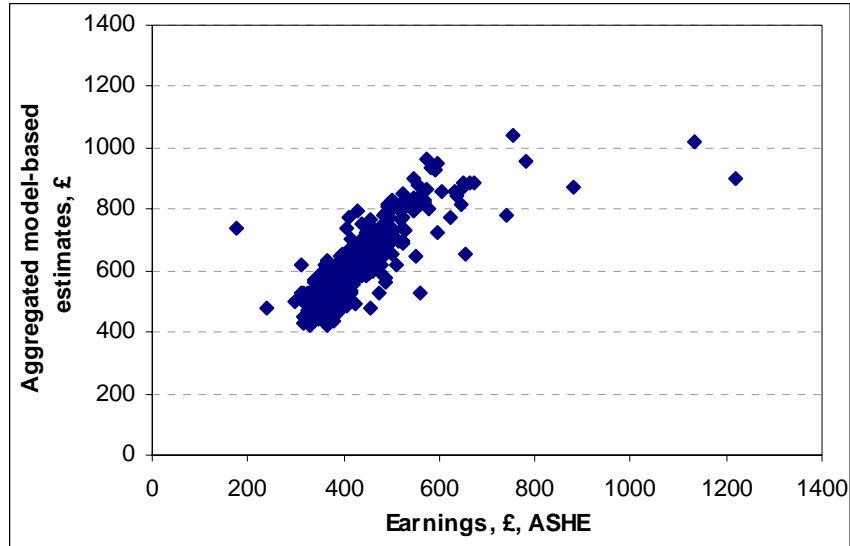


Figure 13: Aggregated model-based estimates compared with ASHE data, England and Wales

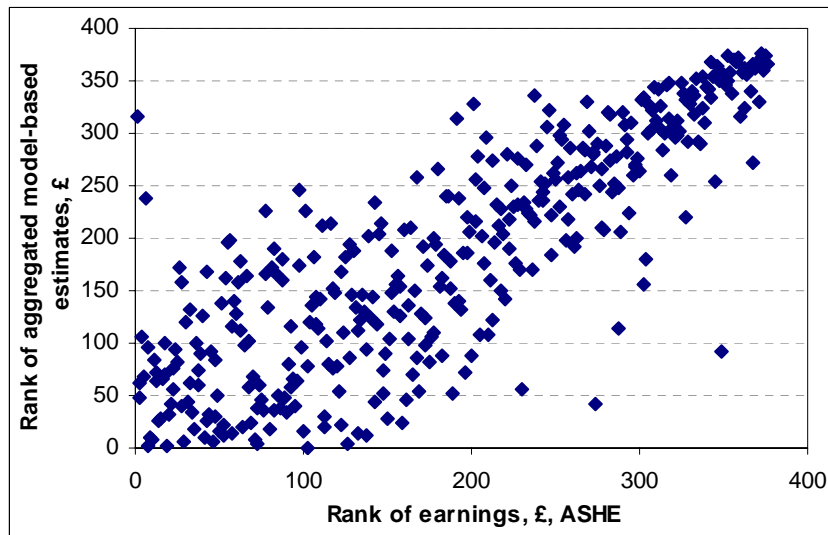


Figure 14: Rank of aggregated model-based estimates compared with rank of ASHE data, England and Wales

Figure 13 shows the aggregated model-based average weekly household income for the majority of LADs are greater than the earnings for a LAD in the ASHE data. This is to be expected, as the model-based income estimates include income from self-employment, investments, benefits and pensions in addition to earnings, which are not included in the ASHE earnings. In addition, the model-based estimates of average

weekly income are calculated for households and the earnings estimates are only for individuals, therefore, household income will be greater than or equal to individual earnings in most cases.

However, Figure 13 shows a small number of LADs where average individual earnings are greater than average weekly household income. These LADs are all located in London. It is likely that this is because these LADs contain MSOAs with both low and high model-based average MSOA incomes. Also the model-based estimates are averaged for households and then over all households in the area, i.e. all private households in the area are included. Whereas, the ASHE survey estimates are calculated as an average over only adults in employment.

Figure 14 compares the ranks of the two data sets and shows a positive relationship, summarised by a rank correlation coefficient of +0.83. Figure 13 and Figure 14 show a wide spread of data points with strong correlation. This pattern is replicated for the other regions. However, the Pearson and Spearman correlation coefficients for the South West are low, 0.40 and 0.70 respectively. This is due to the large difference between the model-based estimate and the ASHE earnings for the LAD Isles of Scilly. If the Isles of Scilly are omitted from the correlation coefficients the figures change to 0.85 and 0.81, hence showing the strong relationship between the two datasets at LAD level. Figure 15 and Figure 16 display the results for London. Here we see the most variability in results.

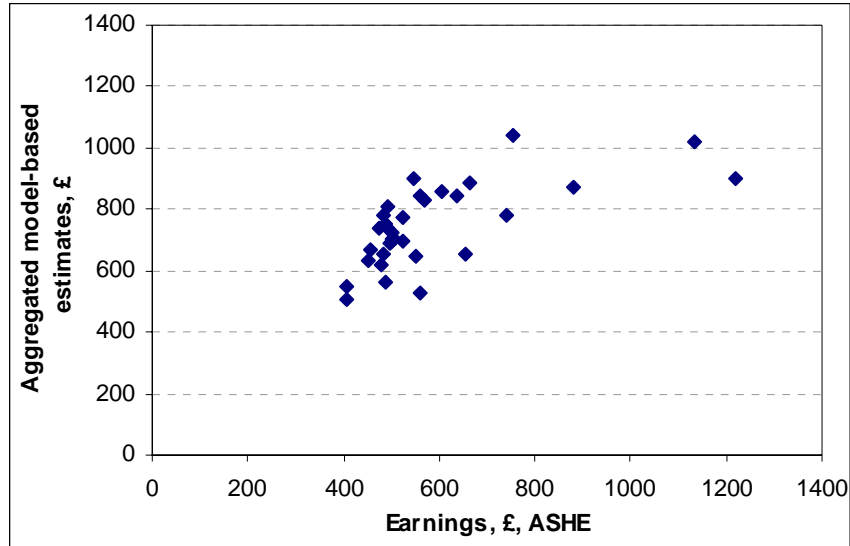


Figure 15: Aggregated model-based estimates compared with ASHE data, London

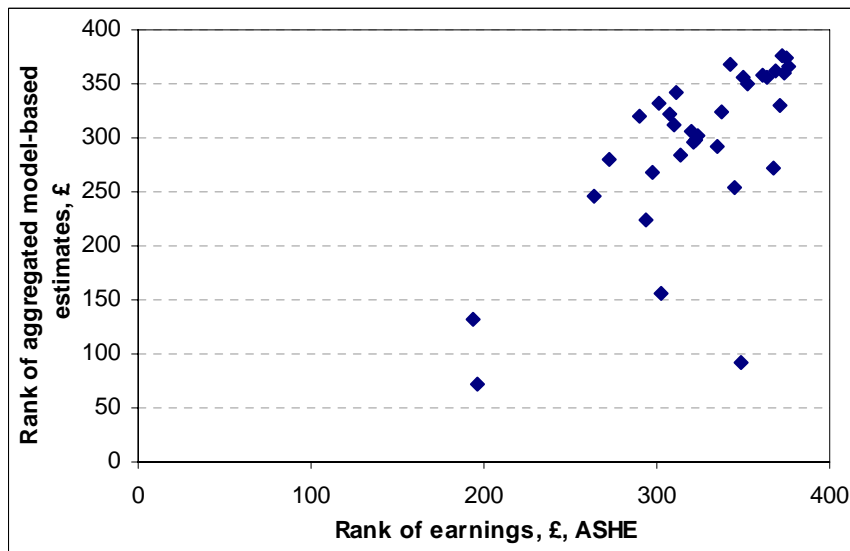


Figure 16: Rank of aggregated model-based estimates compared with rank of ASHE data, London

Figure 15 shows that for some LADs where the LADs contain MSOAs with both low and high model-based average MSOA incomes, e.g. City of London, Westminster and Kensington and Chelsea the average earning value from the ASHE is greater than the aggregated model estimate. However in other LADs such as Harrow and Kingston upon Thames, the aggregated model estimates are much greater than the average

earnings. Figure 16 is very different to the ranking plots for the other regions. The majority of the LADs in London are ranked highly for both the model-based estimates and the ASHE earnings. However, four LADs Newham, Hackney, Tower Hamlets and Barking and Dagenham are ranked highly by average earnings but have a lower ranking on the scale for the model-based estimates of household income. This result is reasonable. Many areas in London will have reasonably high earnings for those employed but the average household income may be very variable.

Appendix A contains a full set of Pearson and Spearman Rank correlations and Appendix B contains a full set of plots for this analysis for total income.

2.2.5 General Household Survey

The General Household Survey (GHS) is another ONS survey that collects some information on income (ONS (2005)). Estimates of average weekly gross **unequalised** household income were obtained from the 2005 survey. These estimates are only available at the GOR level and so the model-based MSOA estimates of average weekly household income were aggregated to this level. The modelling process employed to produce the MSOA-level estimates of income ensures that all MSOA estimates aggregate to the Family Resources Survey (FRS) published estimates at the GOR/country level. Hence in this analysis we are essentially comparing FRS estimates with GHS estimates.

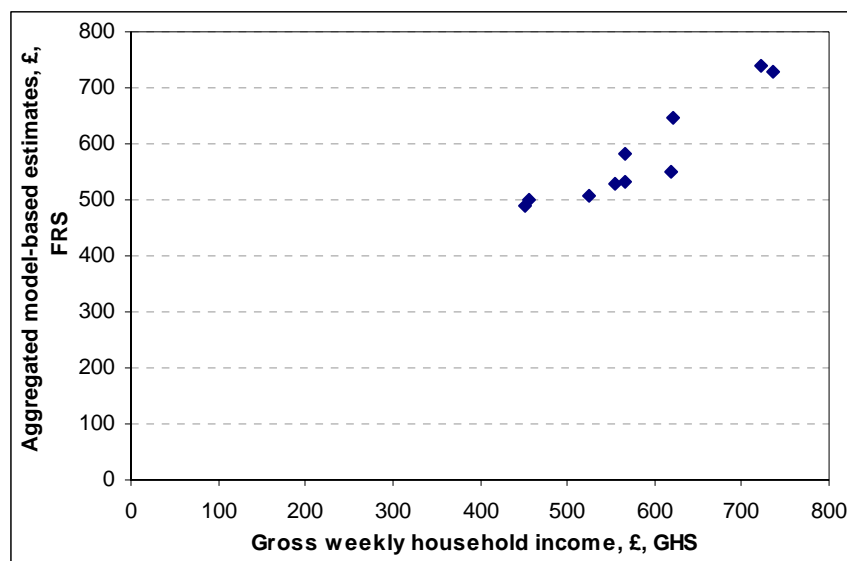


Figure 17: Aggregated model-based estimates compared with GHS data, England and Wales

Figure 17 shows that at the GOR/country level the GHS data and the aggregated model estimates are strongly correlated, summarised by the Pearson correlation coefficient of +0.93.

2.3 Conclusions

MSEA-level estimates of average weekly household income have been produced using a model-based small area estimation approach since no other data sources of household income exist at the MSEA-level. This means that it is not possible to carry out direct validation of the estimates against another data source. However, Section 2.2 contains the results of analyses to compare the model-based estimates with other related sources of income data. These data sources may not exactly describe average weekly household income or may not be at the MSEA-level but are all indicators of levels of income in an area. The data can therefore be used for general comparisons and to establish whether the model-based estimates and the alternative data sources exhibit the same overall patterns of income across the country or region. The following conclusions can be made from comparisons with these alternative income data sources.

- For England and Wales, the correlation between total household income and HM Revenue and Customs total individual income at the LAD level is strong in a positive direction.
- The Index of Multiple Deprivation in England and Wales and the model-based estimates of income rankings have a moderately strong negative relationship. In London (and to a lesser extent in the East and South East GORs) the relationship between the two measures of income are different reflecting the fact that levels of wealth within individual MSOAs can be very heterogeneous in these areas.
- For the Annual Survey of Hours and Earnings which provides estimates of earnings in LADs the correlation between earnings and household income is strong in a positive direction.
- At the GOR/country level the GHS estimates of gross weekly **unequalised** household income are strongly correlated with the aggregated model-based estimates.

Although it is difficult to draw strong conclusions from these analyses, overall the results are consistent across the different income types and exhibit the expected patterns. In most cases the correlation between the model-based estimates and alternative income data sources are moderately strong. Many outliers have been explained. In particular different patterns have been observed in London, as expected, due to the local variability of income levels.

Appendix

A Statistical Results

This appendix contains a full set of statistical results for the analyses described in Section 2.2 for average total weekly household income. The Pearson correlation coefficient and the Spearman Rank correlation coefficient are displayed for the full data set as well as for individual regions (if appropriate) for each different income data source.

A.1 Her Majesty's Revenue and Customs data – England and Wales

	Pearson correlation coefficient	Spearman rank correlation coefficient
England and Wales	0.76	0.92
North East	0.89	0.78
North West	0.87	0.84
Yorkshire and The Humber	0.94	0.95
East Midlands	0.93	0.90
West Midlands	0.91	0.89
East	0.93	0.92
London	0.64	0.76
South East	0.93	0.95
South West	0.87	0.89
Wales	0.90	0.80

A.2 Index of Multiple Deprivation – England

	Pearson correlation coefficient	Spearman rank correlation coefficient
England	-0.58	-0.70
North East	-0.78	-0.77
North West	-0.82	-0.85
Yorkshire and The Humber	-0.87	-0.86
East Midlands	-0.87	-0.92
West Midlands	-0.83	-0.82
East	-0.79	-0.79
London	-0.77	-0.78
South East	-0.88	-0.91
South West	-0.83	-0.84

A.3 Index of Multiple Deprivation – Wales

	Pearson correlation coefficient	Spearman rank correlation coefficient
Wales	-0.64	-0.55

A.4 Annual Survey of Hours and Earnings – England and Wales

	Pearson correlation coefficient	Spearman rank correlation coefficient
England and Wales	0.79	0.83
North East	0.61	0.66
North West	0.83	0.77
Yorkshire and The Humber	0.59	0.56
East Midlands	0.85	0.86
West Midlands	0.80	0.76
East of England	0.91	0.92
London	0.66	0.69
South East	0.88	0.91
South West	0.40	0.70
Wales	0.79	0.71

A.5 General Household Survey – England, Wales, Scotland and Northern Ireland

	Pearson correlation coefficient
England and Wales	0.93

B Graphical Results

This appendix contains a full set of graphical results for the analyses described in Section 2.2 for average total weekly household income. The plots for the full data set as well as regional plots (if appropriate) are displayed for each different income data source.

B.1 HM Revenue and Customs data – England and Wales

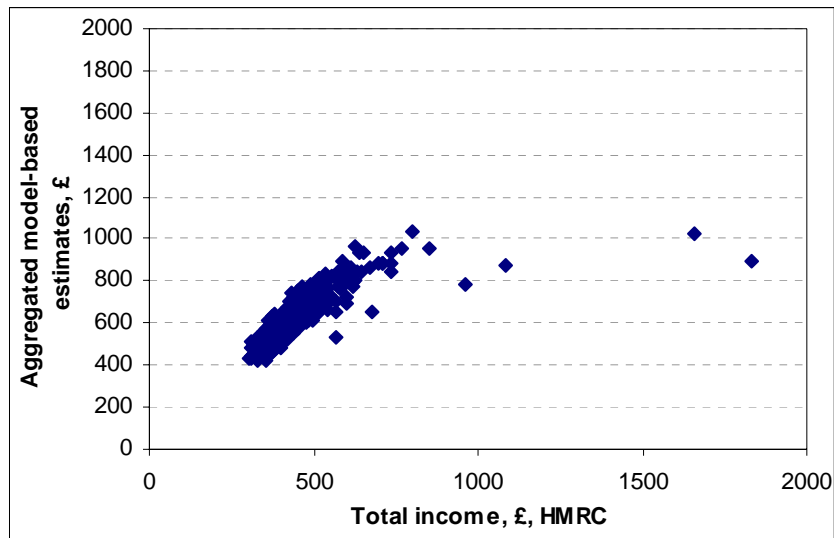


Figure 18: Aggregated model-based estimates compared with HMRC data, England and Wales

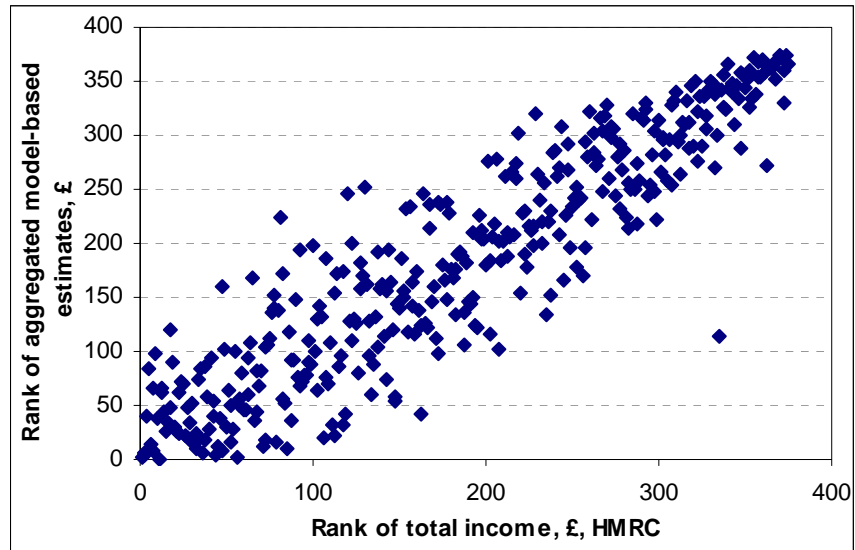


Figure 19: Rank of aggregated model-based estimates compared with HMRC data, England and Wales

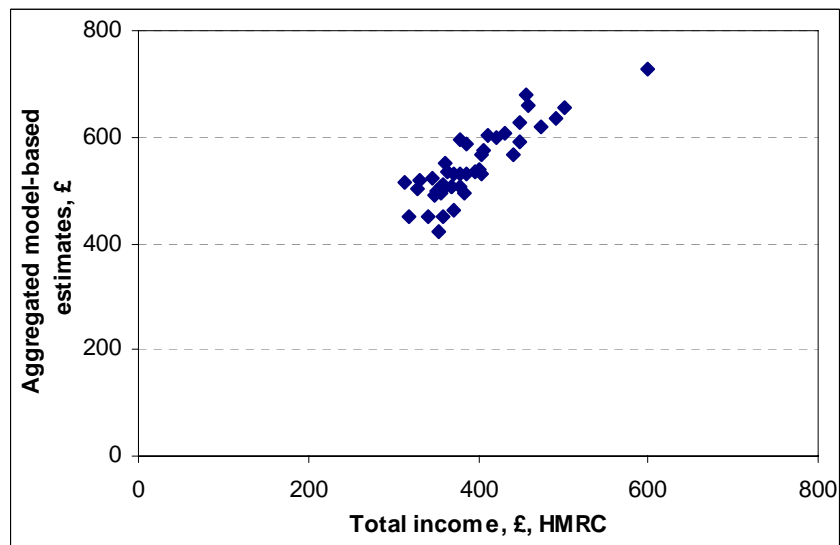


Figure 20: Aggregated model-based estimates compared with HMRC data, North West

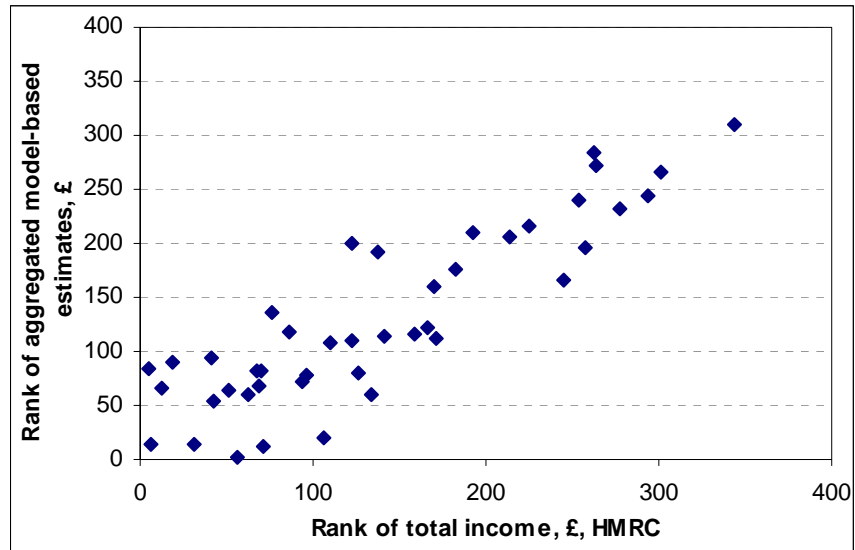


Figure 21: Rank of aggregated model-based estimates compared with rank of HMRC data, North West

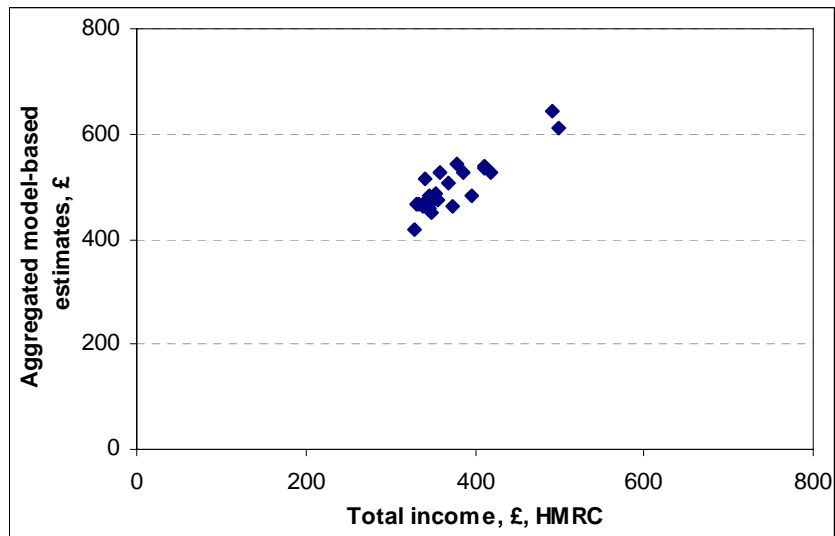


Figure 22: Aggregated model-based estimates compared with HMRC data, North East

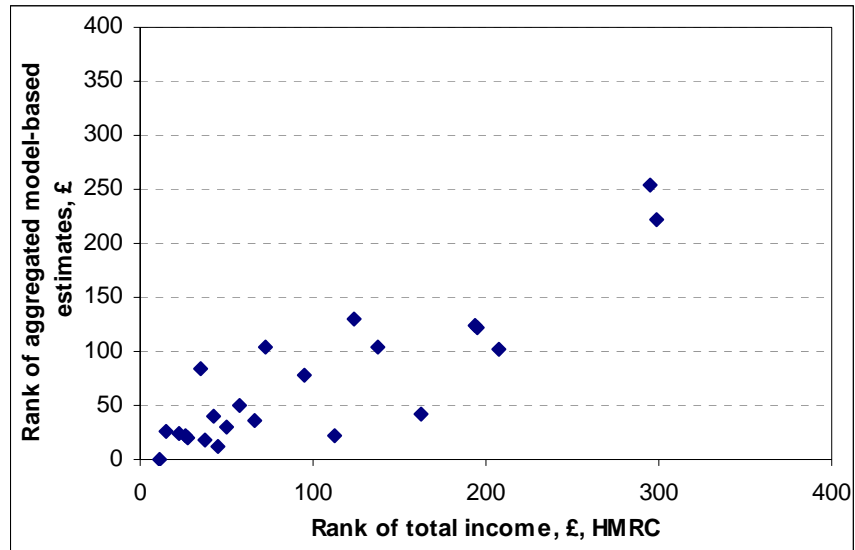


Figure 23: Rank of aggregated model-based estimates compared with rank of HMRC data, North East

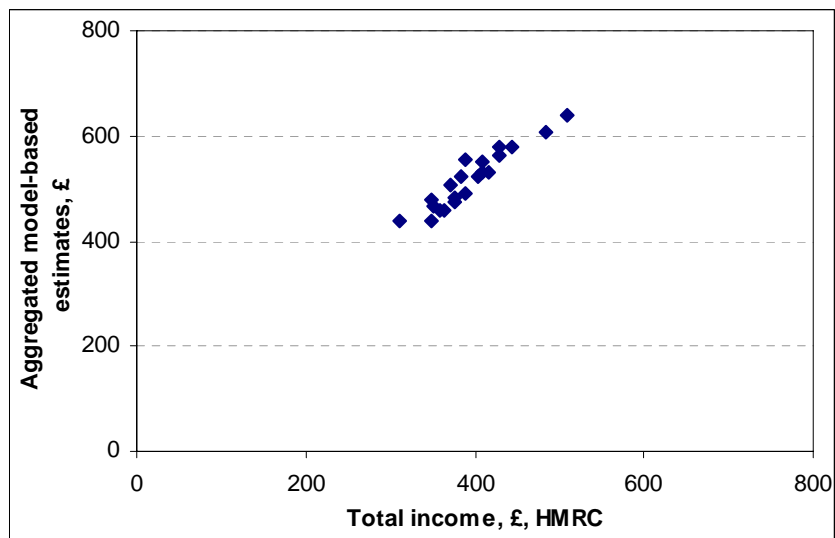


Figure 24: Aggregated model-based estimates compared with HMRC data, Yorkshire and The Humber

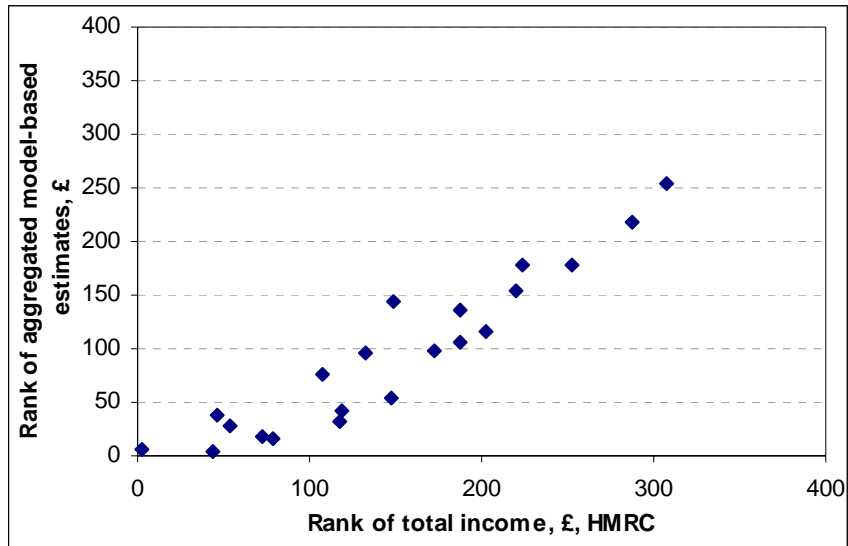


Figure 25: Rank of aggregated model-based estimates compared with rank of HMRC data, Yorkshire and The Humber

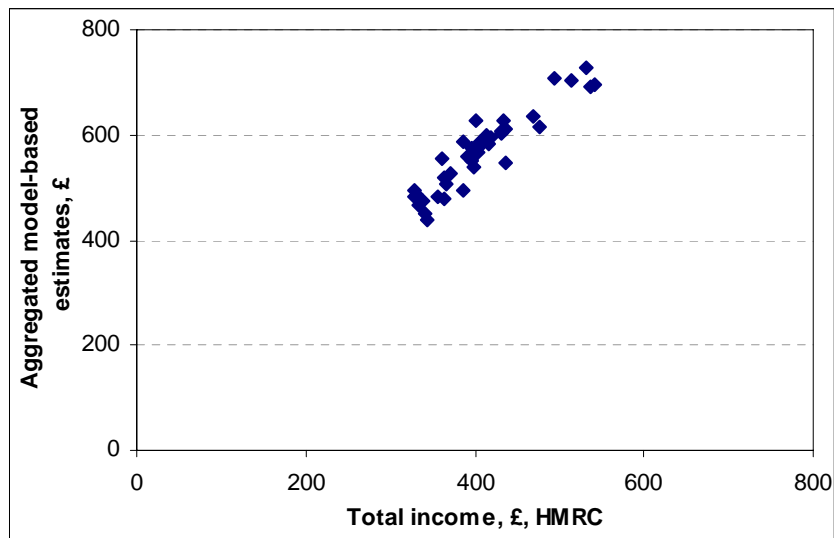


Figure 26: Aggregated model-based estimates compared with HMRC data, East Midlands

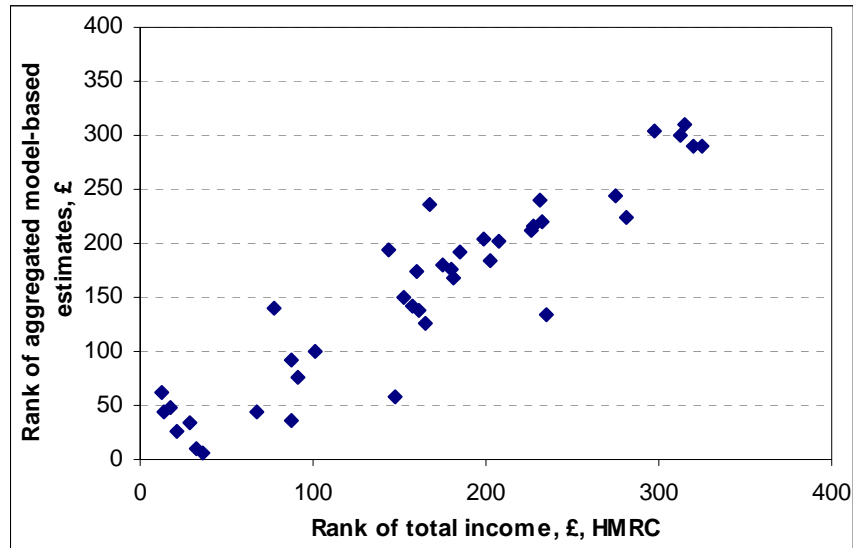


Figure 27: Rank of aggregated model-based estimates compared with rank of HMRC data, East Midlands

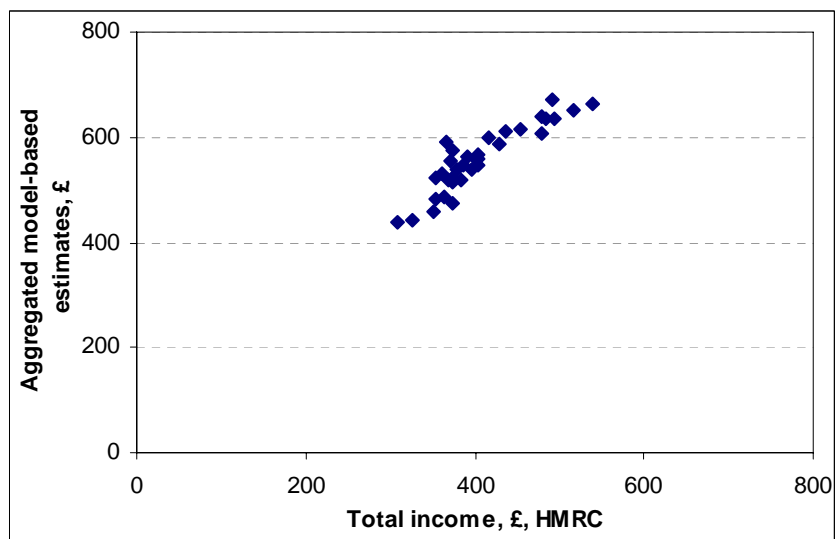


Figure 28: Aggregated model-based estimates compared with HMRC data, West Midlands

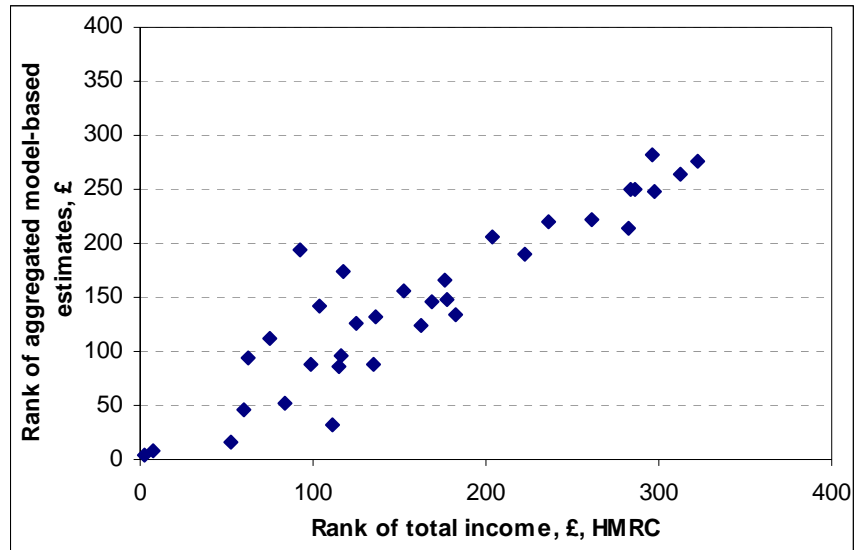


Figure 29: Rank of aggregated model-based estimates compared with rank of HMRC data, West Midlands

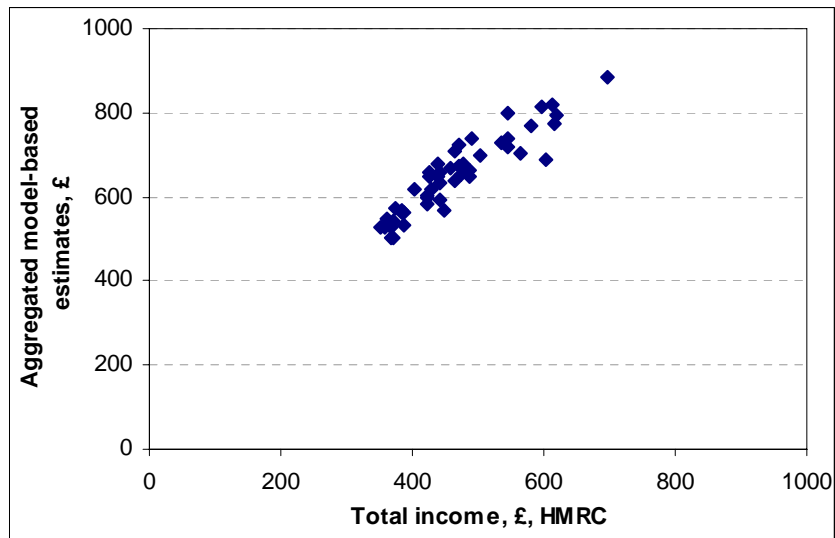


Figure 30: Aggregated model-based estimates compared with HMRC data, East

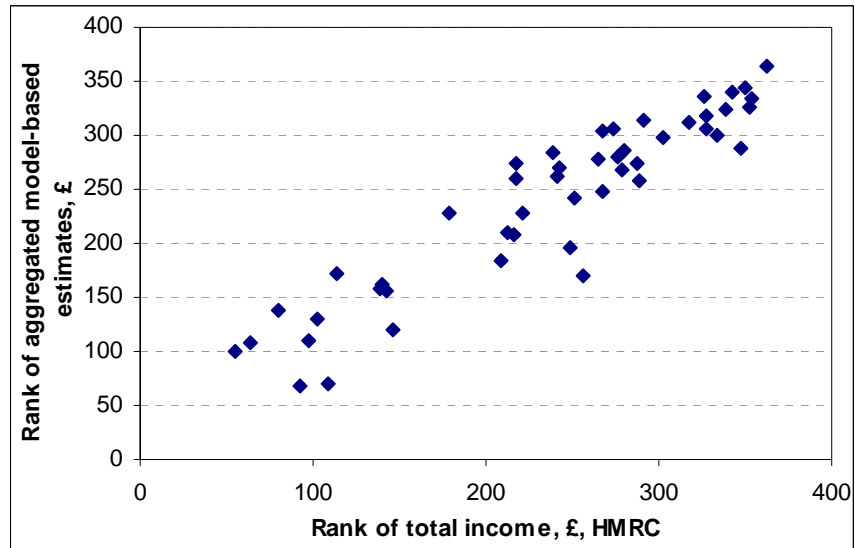


Figure 31: Rank of aggregated model-based estimates compared with rank of HMRC data, East

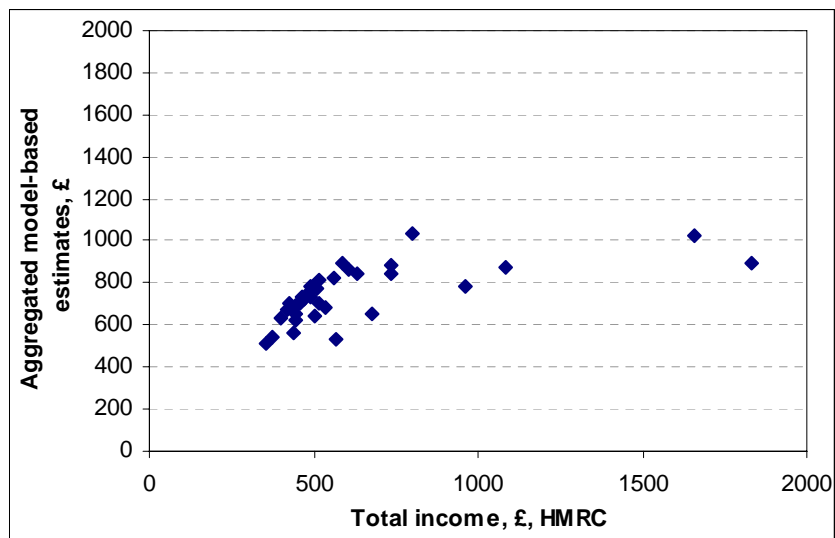


Figure 32: Aggregated model-based estimates compared with HMRC data, London

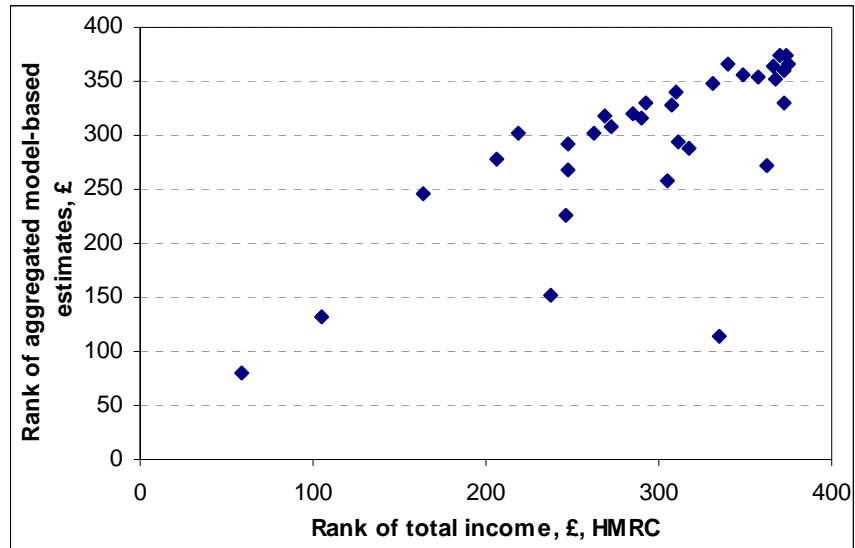


Figure 33: Rank of aggregated model-based estimates compared with rank of HMRC data, London

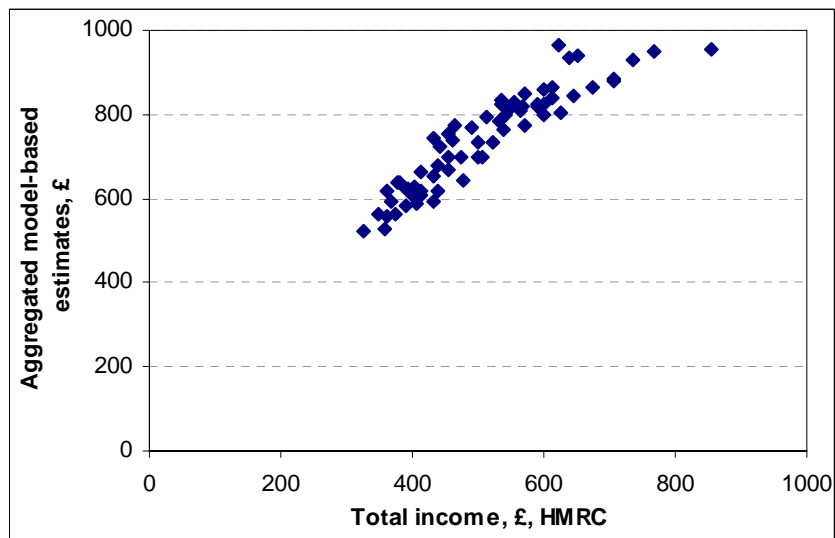


Figure 34: Aggregated model-based estimates compared with HMRC data, South East

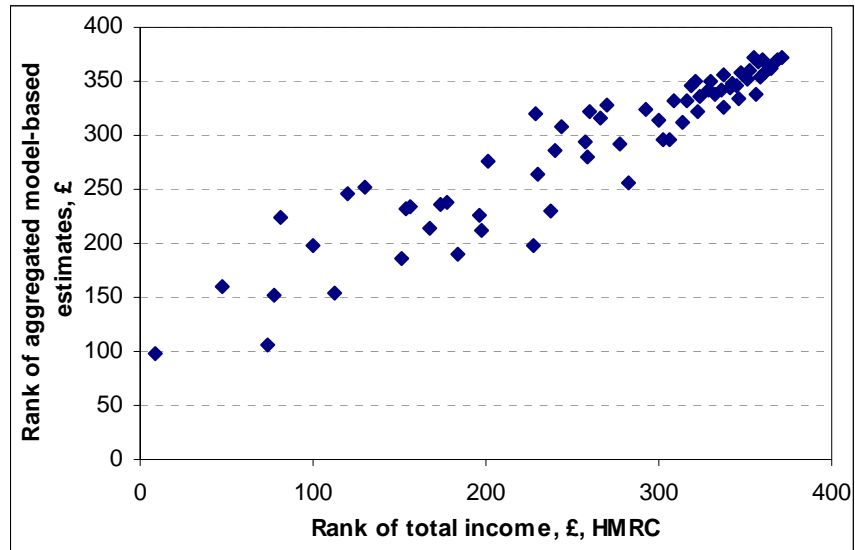


Figure 35: Rank of aggregated model-based estimates compared with HMRC data,
South East

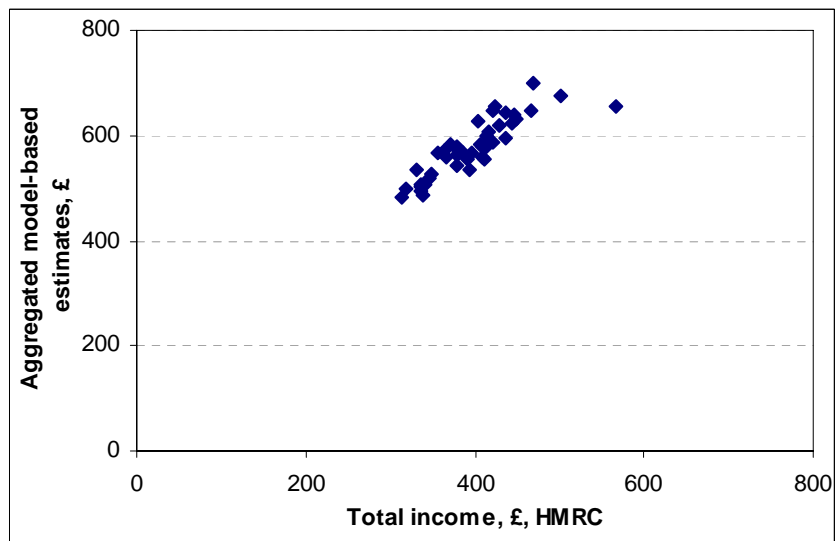


Figure 36: Aggregated model-based estimates compared with HMRC data, South
West

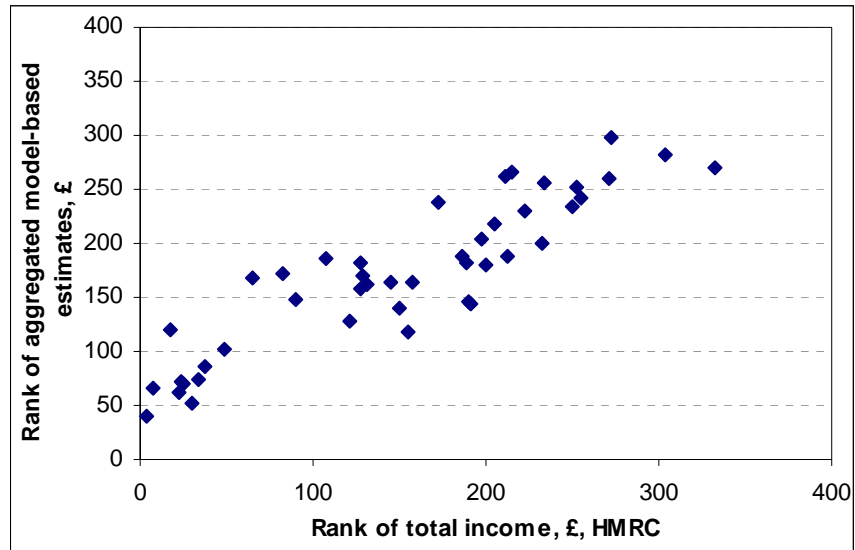


Figure 37: Rank of aggregated model-based estimates compared with rank of HMRC data, South West

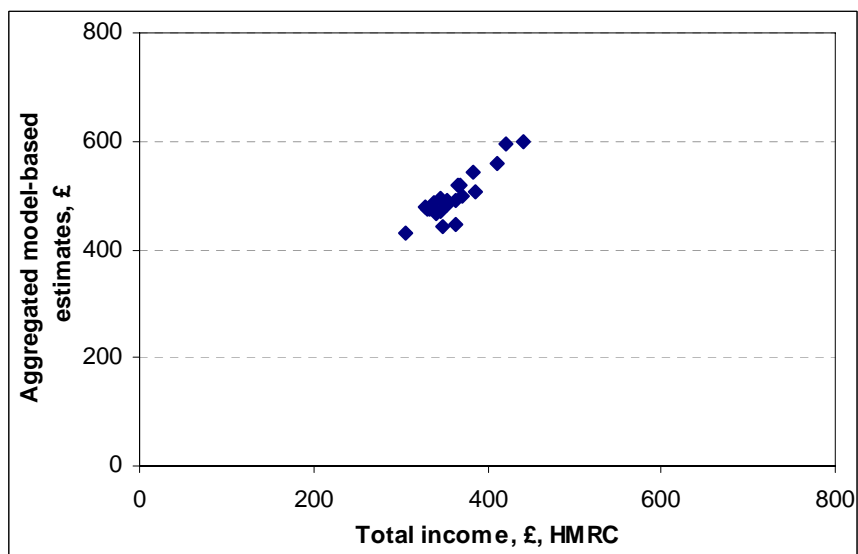


Figure 38: Aggregated model-based estimates compared with HMRC data, Wales

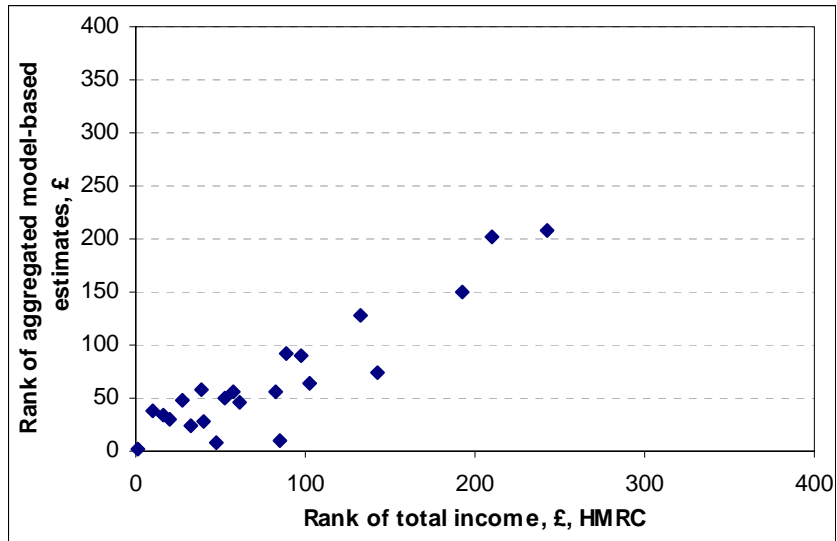


Figure 39: Rank of aggregated model-based estimates compared with rank of HMRC data, Wales

B.2 Index of Multiple Deprivation – England

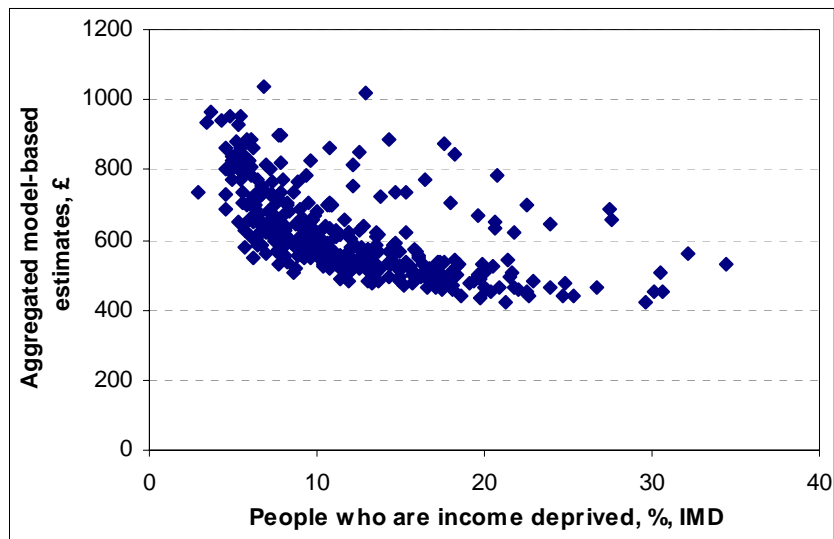


Figure 40: Aggregated model-based estimates compared with Income Score, IMD, England



Figure 41: Rank of aggregated model-based estimates compared with rank of Income Score, IMD, England

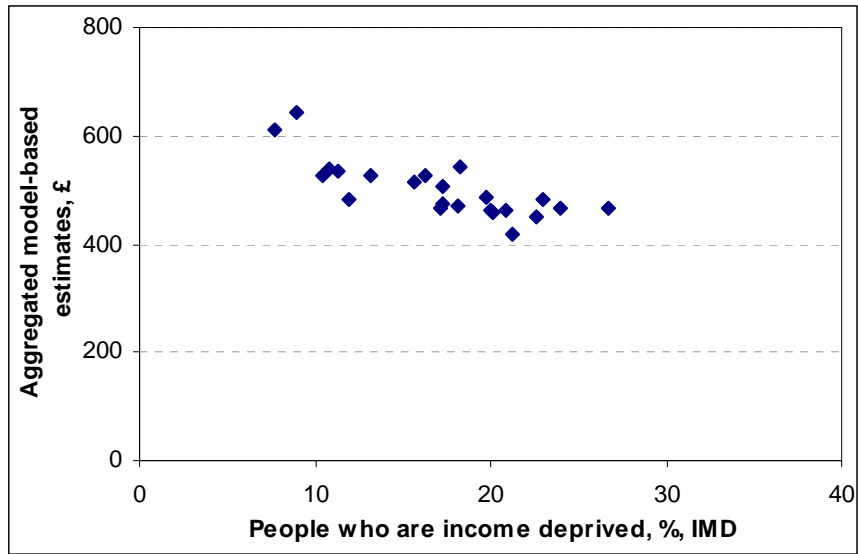


Figure 42: Aggregated model-based estimates compared with Income Score IMD, North East

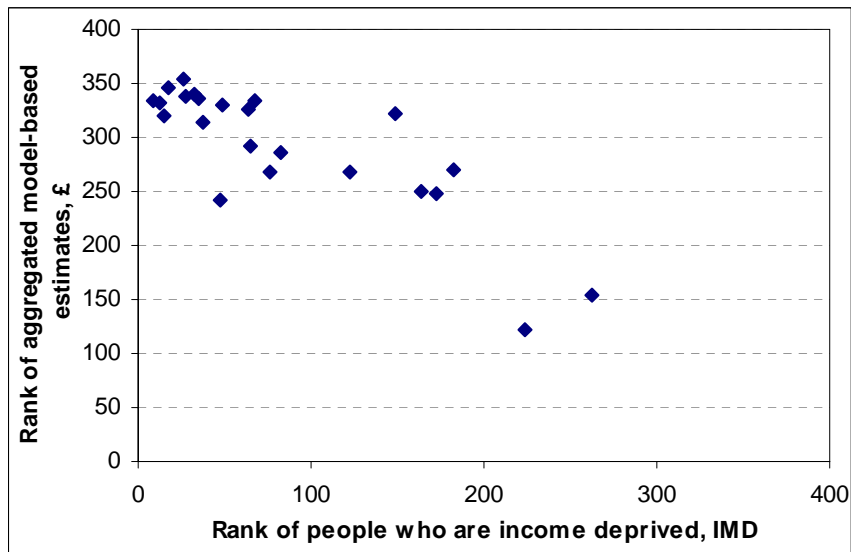


Figure 43: Rank of aggregated model-based estimates compared with rank of Income IMD, North East

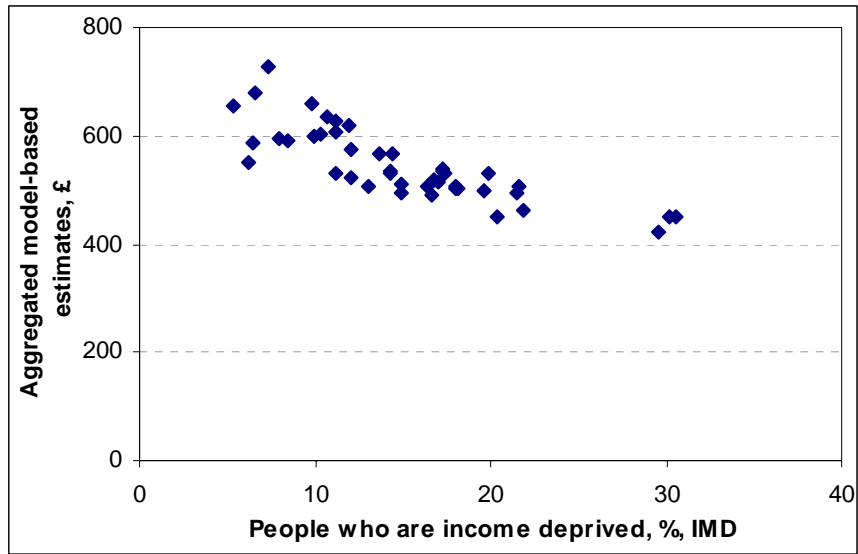


Figure 44: Aggregated model-based estimates compared with Income Score, IMD, North West

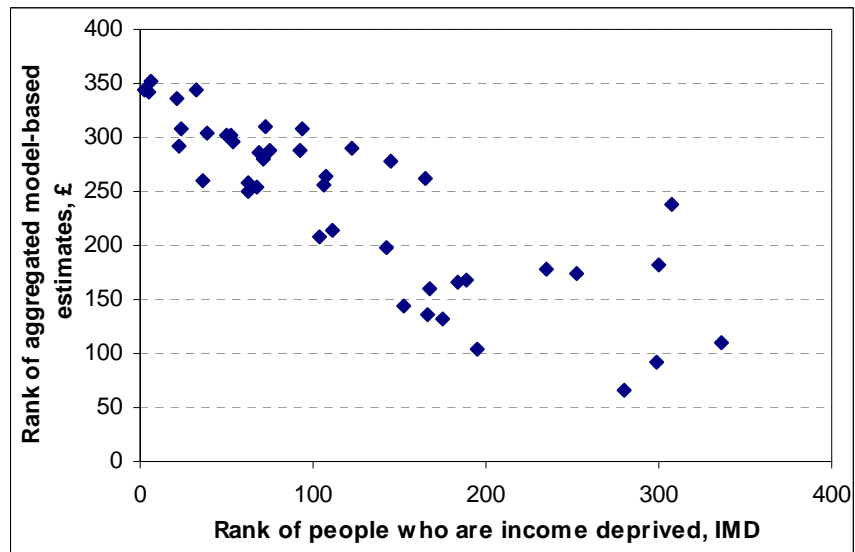


Figure 45: Rank of aggregated model-based estimates compared with rank of Income Score, IMD, North West

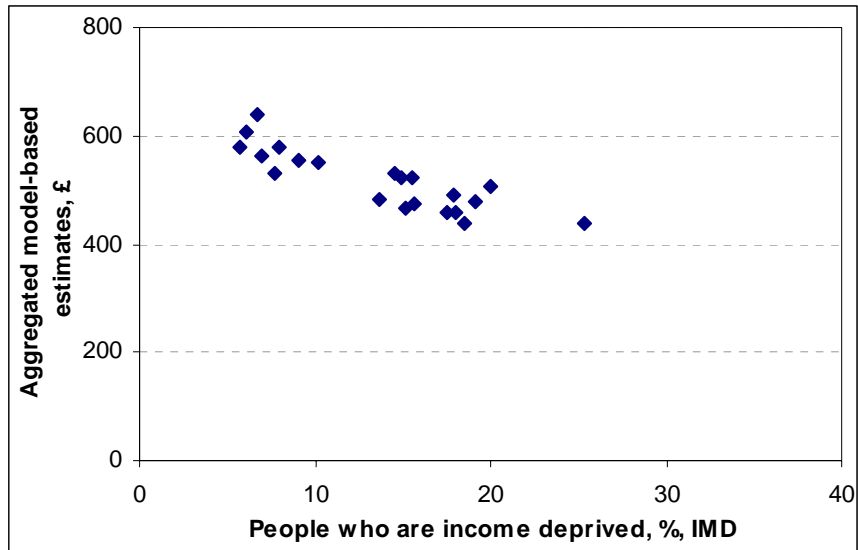


Figure 46: Aggregated model-based estimates compared with Income Score, IMD, Yorkshire and The Humber

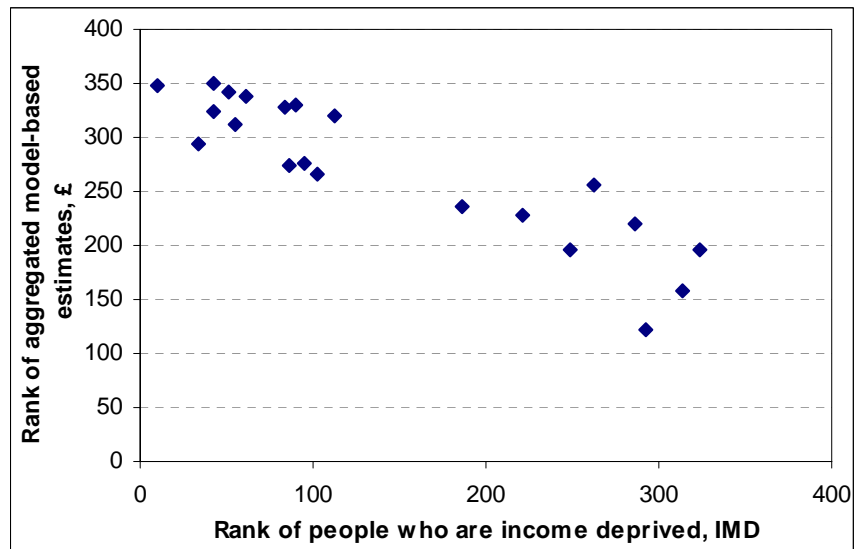


Figure 47: Rank of aggregated model-based estimates compared with rank of Income Score, IMD, Yorkshire and The Humber

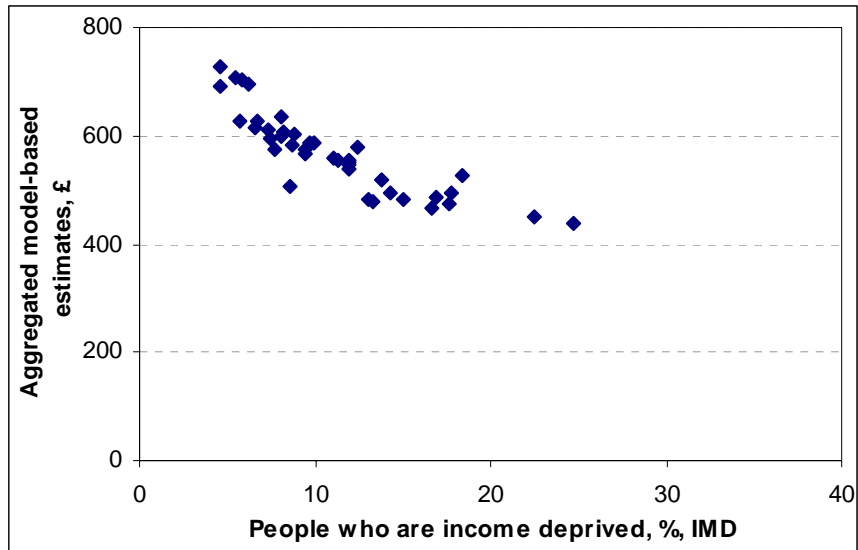


Figure 48: Aggregated model-based estimates compared with Income Score, IMD, East Midlands

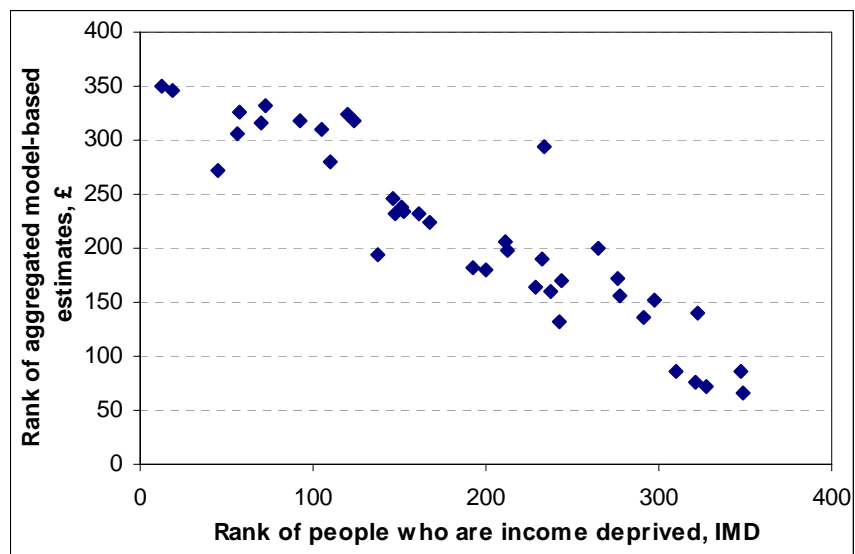


Figure 49: Rank of aggregated model-based estimates compared with rank of Income IMD, East Midlands

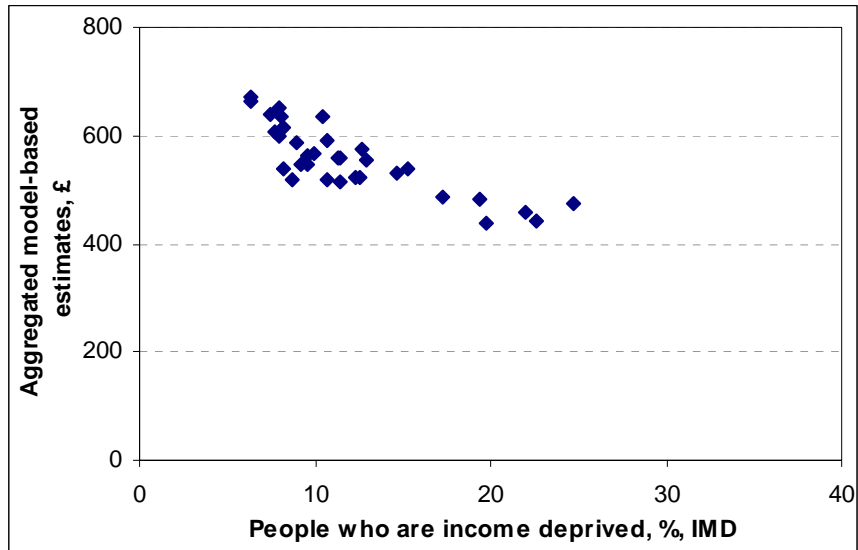


Figure 50: Aggregated model-based estimates compared with Income Score IMD,
West Midlands

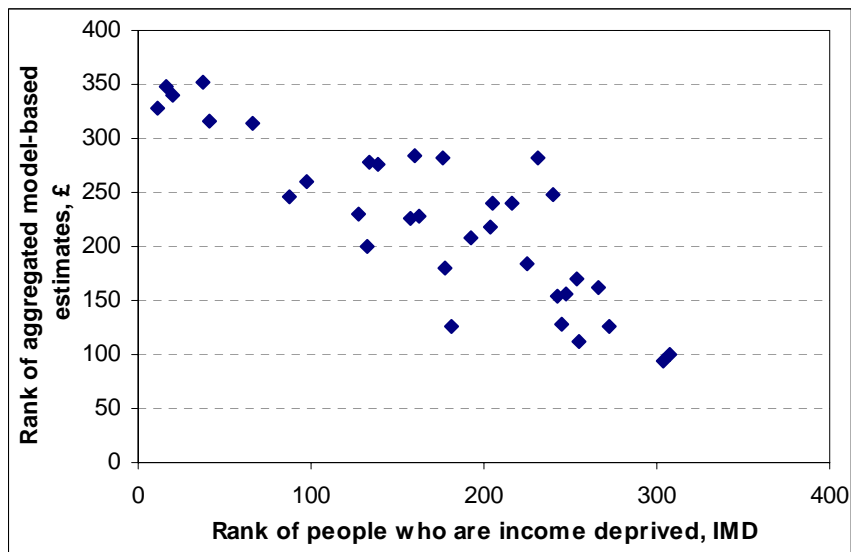


Figure 51: Rank of aggregated model-based estimates compared with rank of Income
Score, IMD, West Midlands

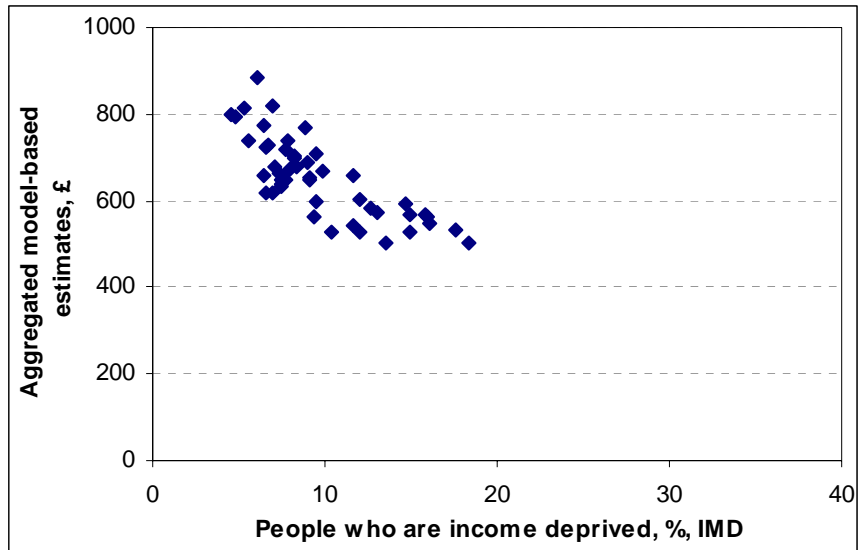


Figure 52: Aggregated model-based estimates compared with Income Score, IMD, East

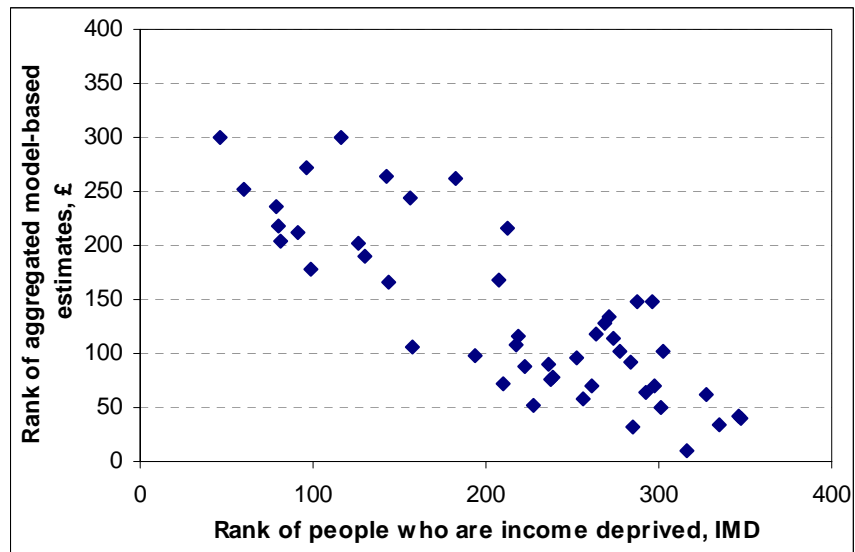


Figure 53: Rank of aggregated model-based estimates compared with rank of Income Score, IMD, East

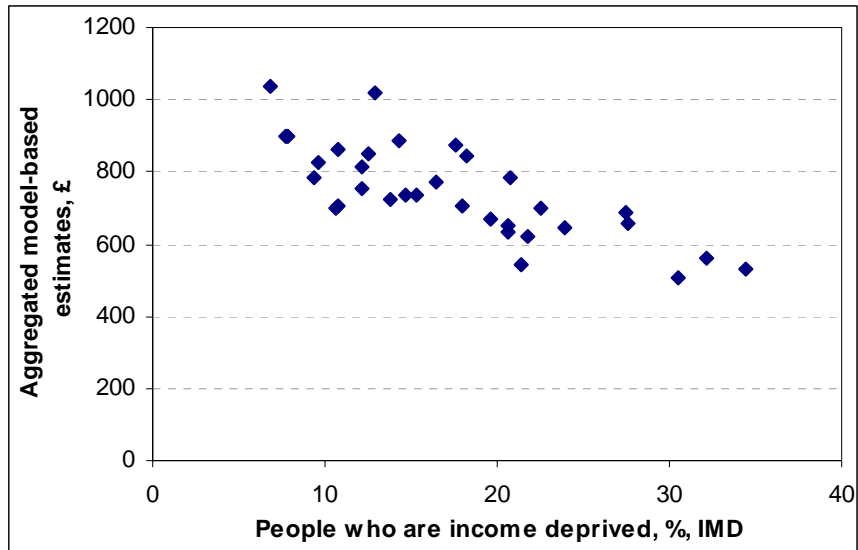


Figure 54: Aggregated model-based estimates compared with Income Score, IMD, London

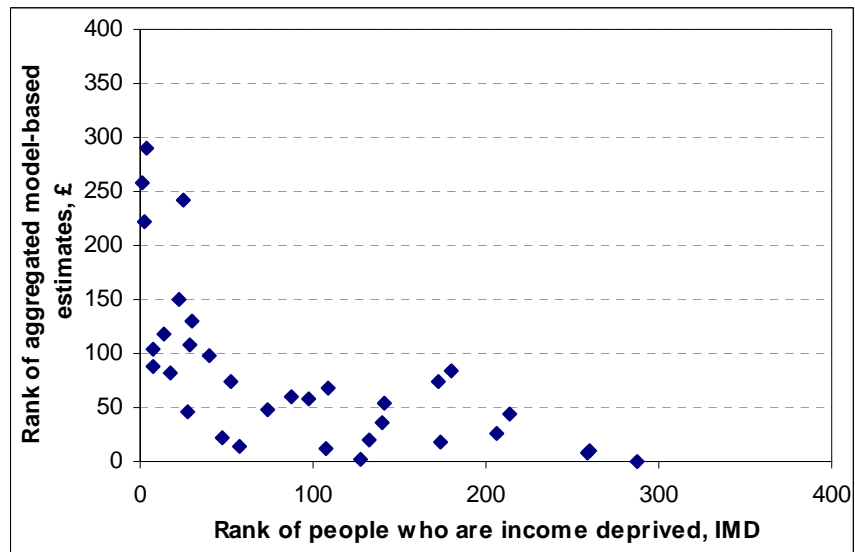


Figure 55: Rank of aggregated model-based estimates compared with rank of Income Score, IMD, London

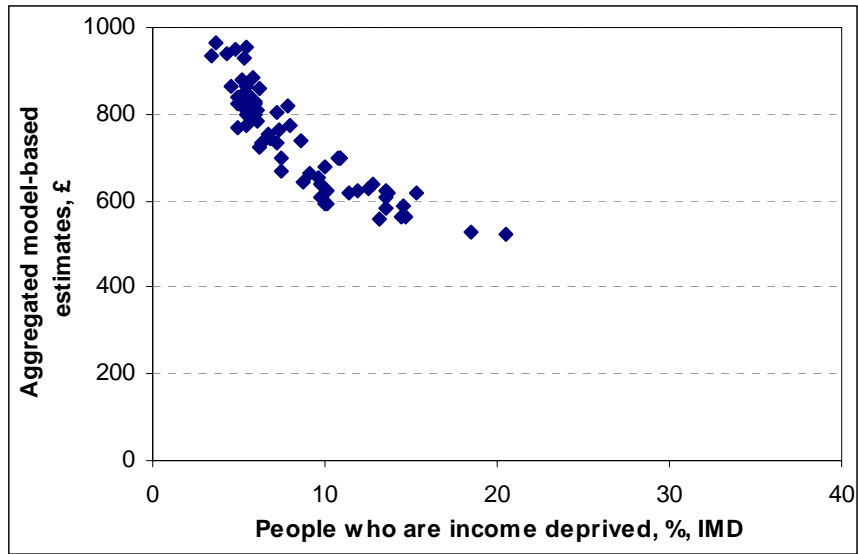


Figure 56: Aggregated model-based estimates compared with Income Score, IMD, South East

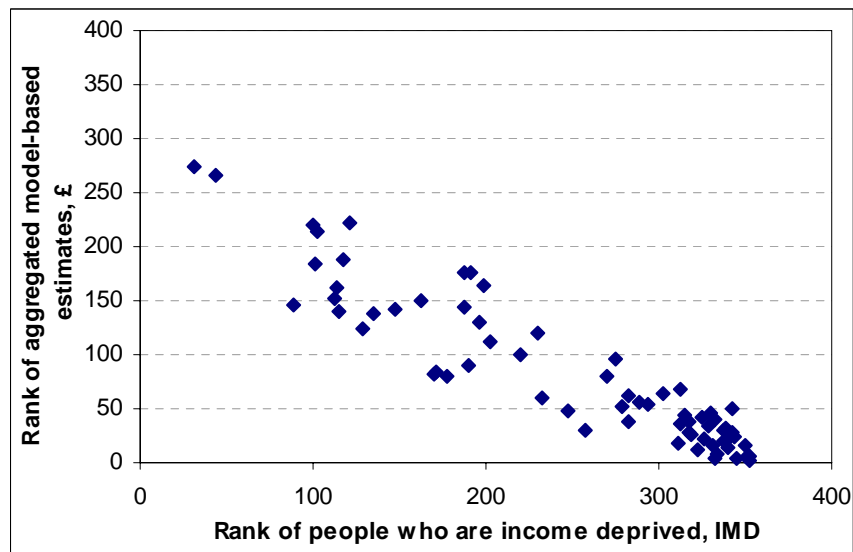


Figure 57: Rank of aggregated model-based estimates compared with rank of Income Score, IMD, South East

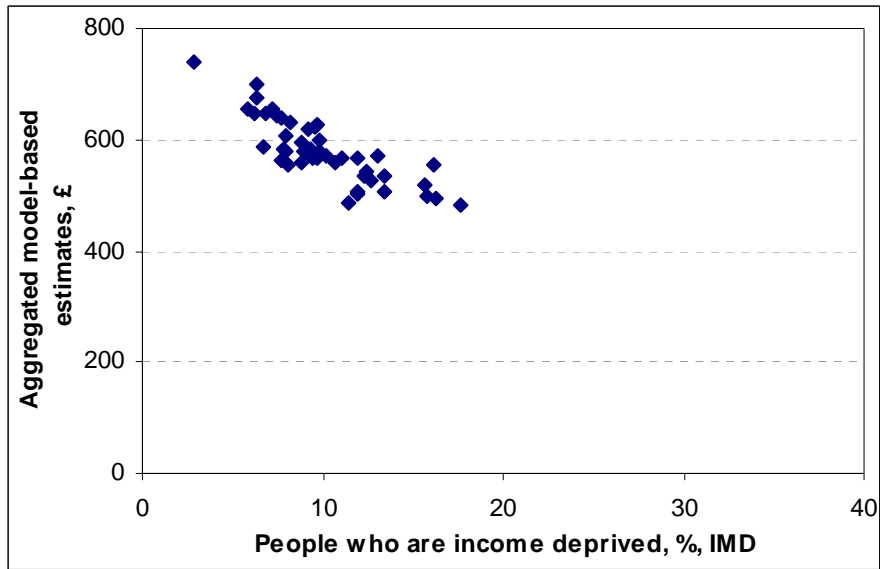


Figure 58: Aggregated model-based estimates compared with Income Score, IMD, South West

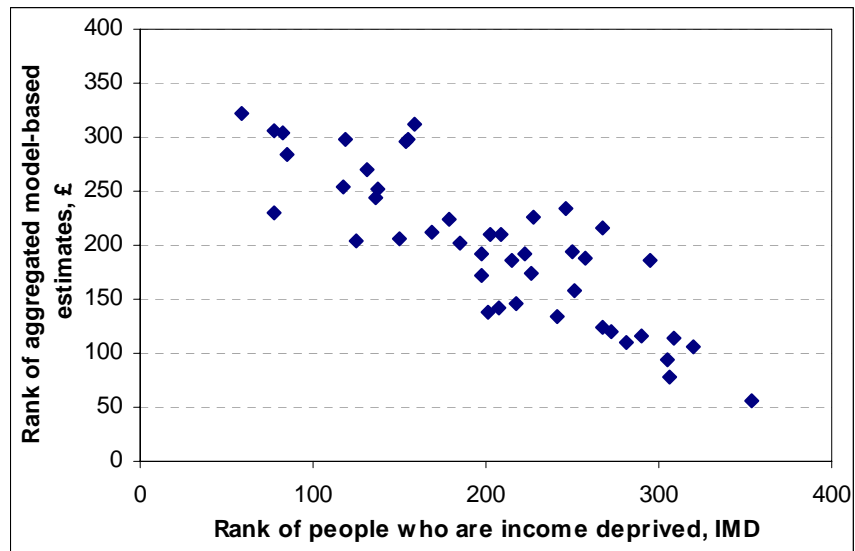


Figure 59: Rank of aggregated model-based estimates compared with rank of Income Score, IMD, South West

B.3 Index of Multiple Deprivation – Wales

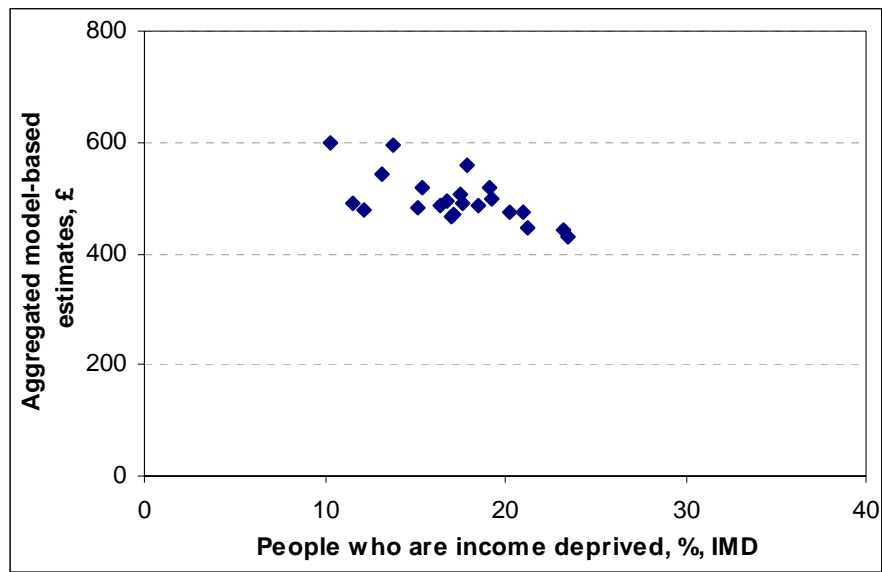


Figure 60: Aggregated model-based estimates compared with Income Score, IMD, Wales

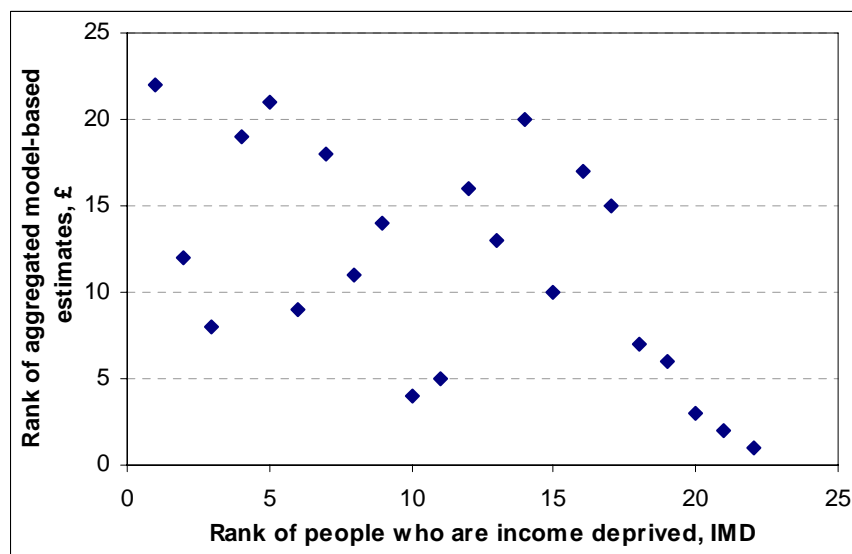


Figure 61: Rank of Aggregated model-based estimates compared with rank of Income Score, IMD, Wales

B.4 Annual Survey of Hours and Earnings – England and Wales

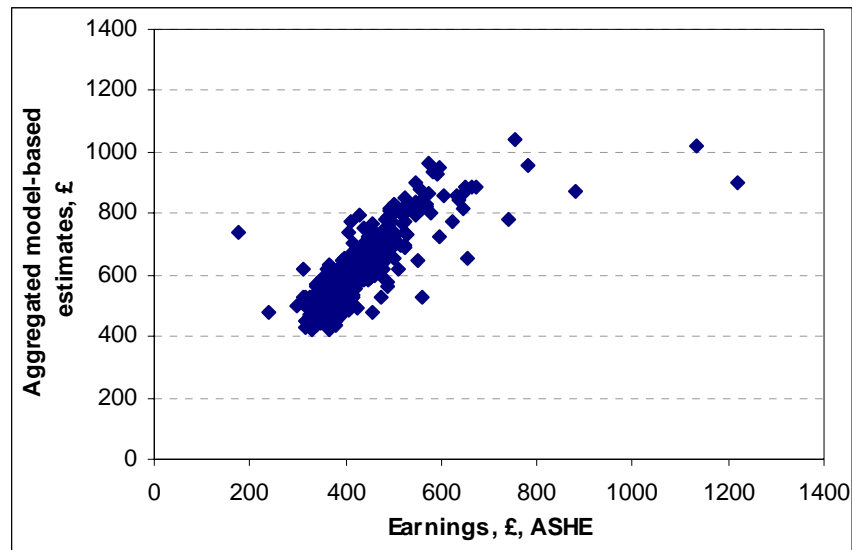


Figure 62: Aggregated model-based estimates compared with ASHE data, England and Wales

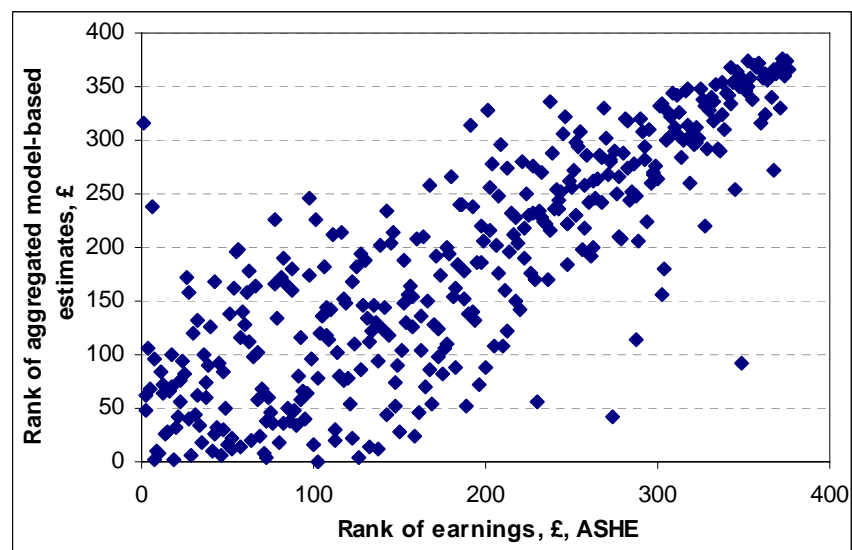


Figure 63: Rank of aggregated model-based estimates compared with rank of ASHE data, England and Wales

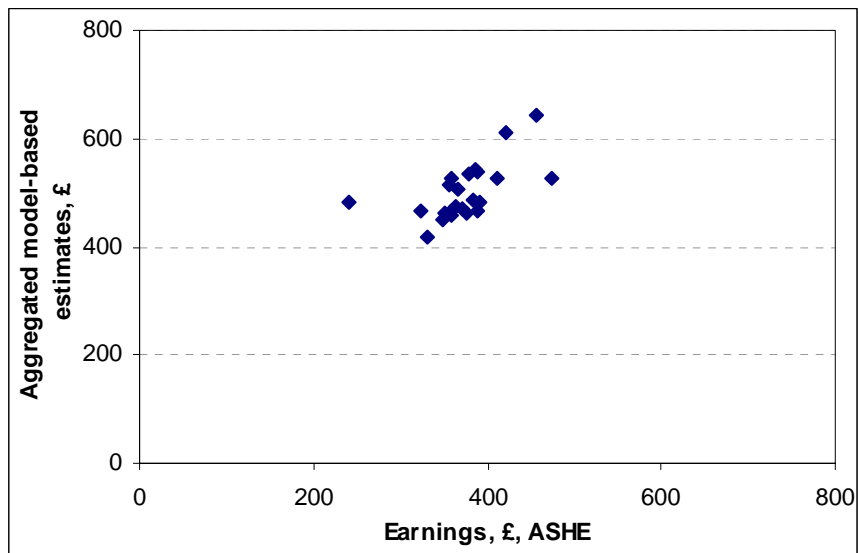


Figure 64: Aggregated model-based estimates compared with ASHE data, North East

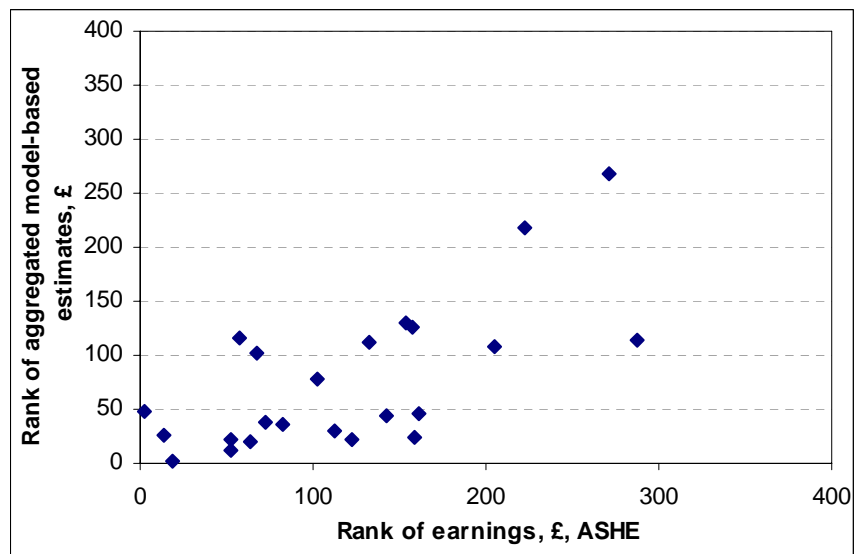


Figure 65: Rank of aggregated model-based estimates compared with rank of ASHE data, North East

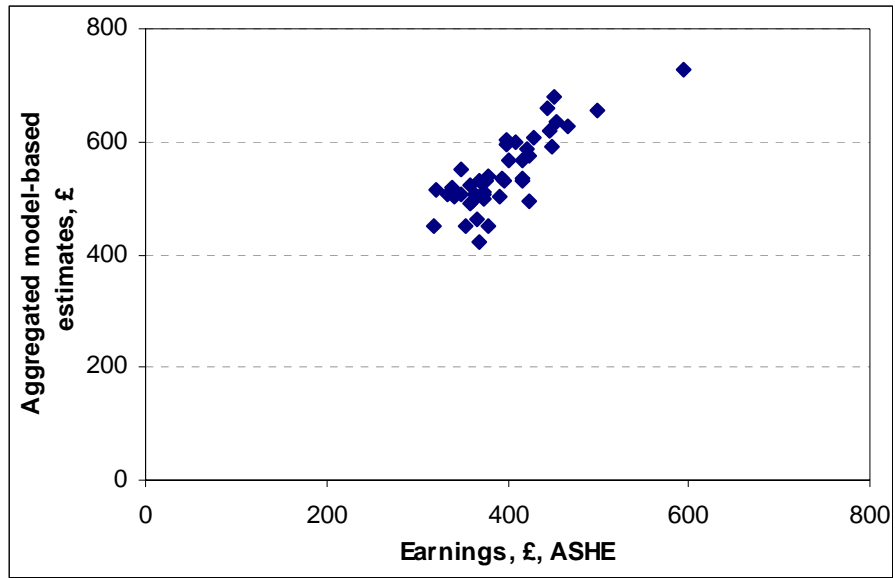


Figure 66: Aggregated model-based estimates compared with ASHE data, North West

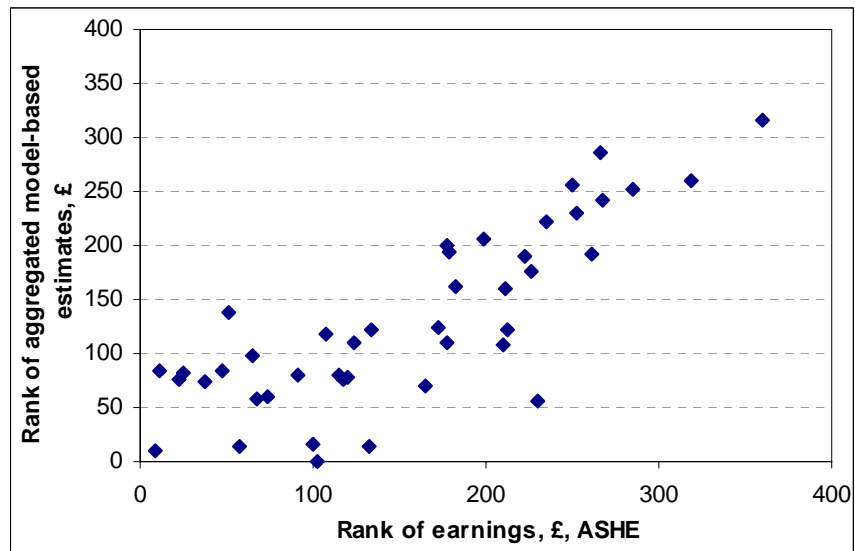


Figure 67: Rank of aggregated model-based estimates compared with rank of ASHE data, North West

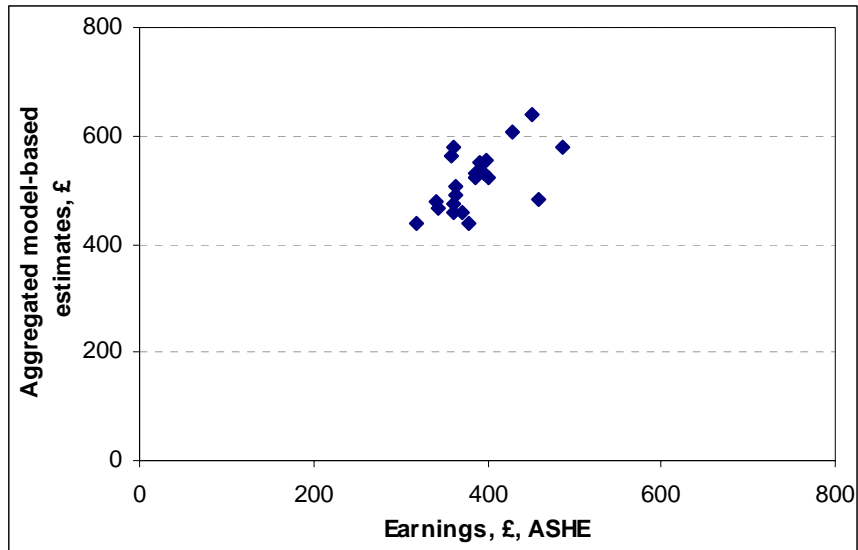


Figure 68: Aggregated model-based estimates compared with ASHE data, Yorkshire and The Humber

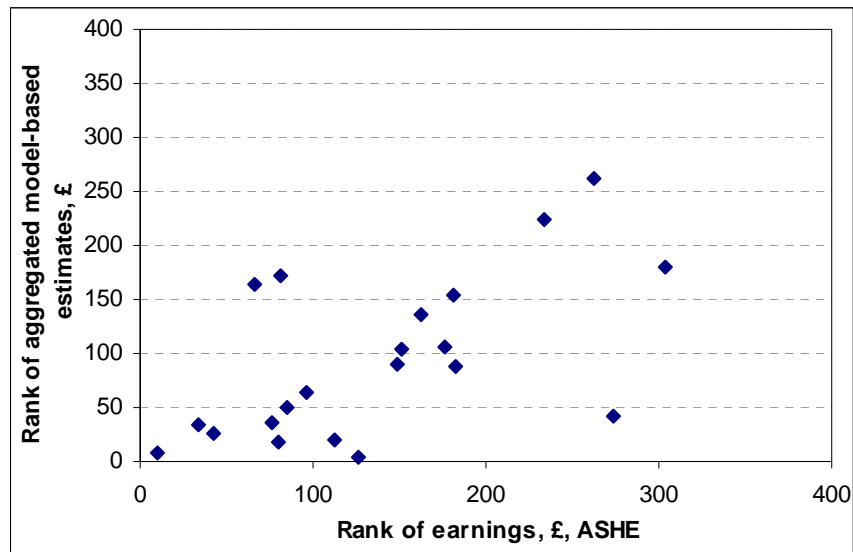


Figure 69: Rank of aggregated model-based estimates compared with rank of ASHE data, Yorkshire and The Humber

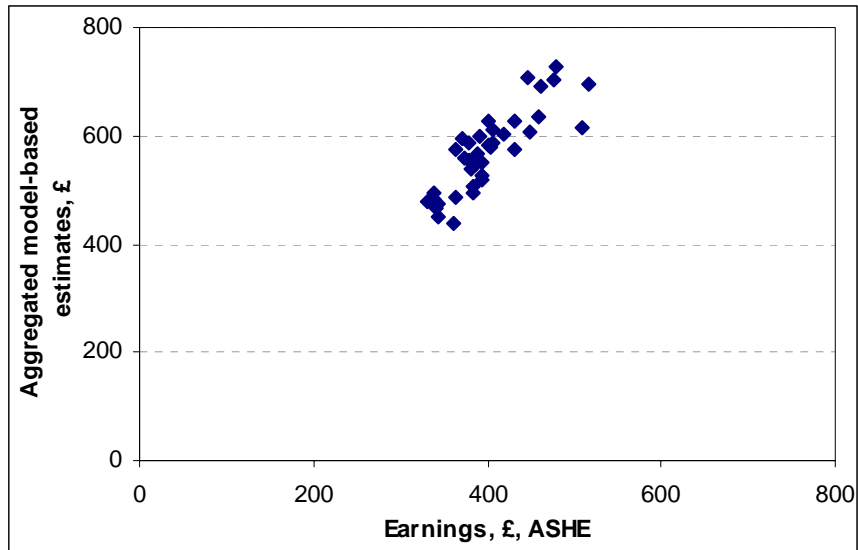


Figure 70: Aggregated model-based estimates compared with ASHE data, East Midlands

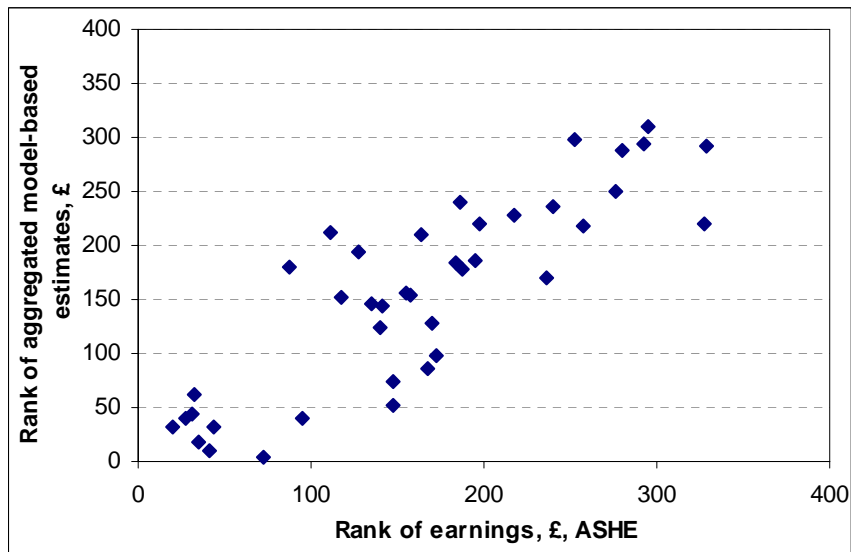


Figure 71: Rank of aggregated model-based estimates compared with rank of ASHE data, East Midlands

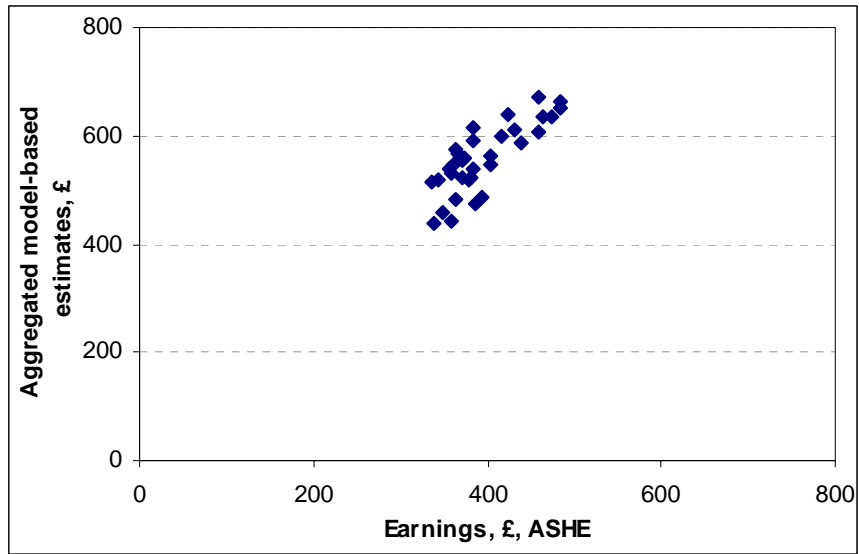


Figure 72: Aggregated model-based estimates compared with ASHE data, West Midlands

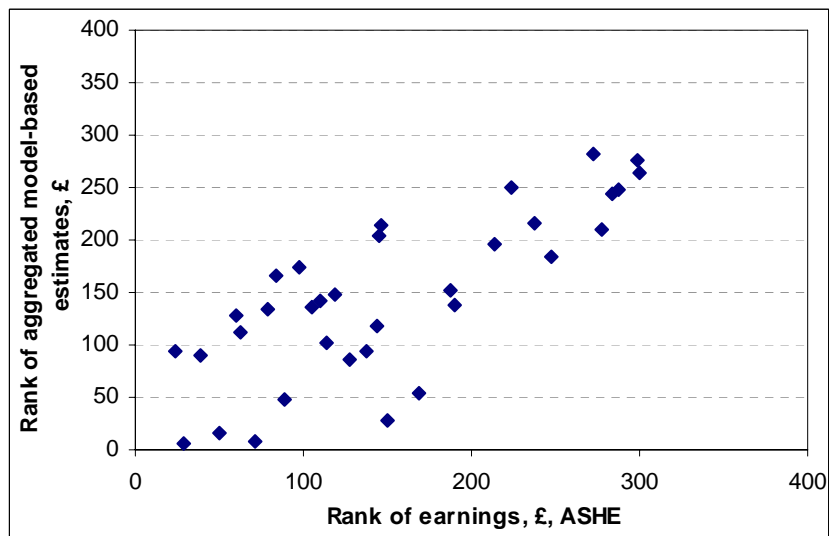


Figure 73: Rank of aggregated model-based estimates compared with rank of ASHE data, West Midlands

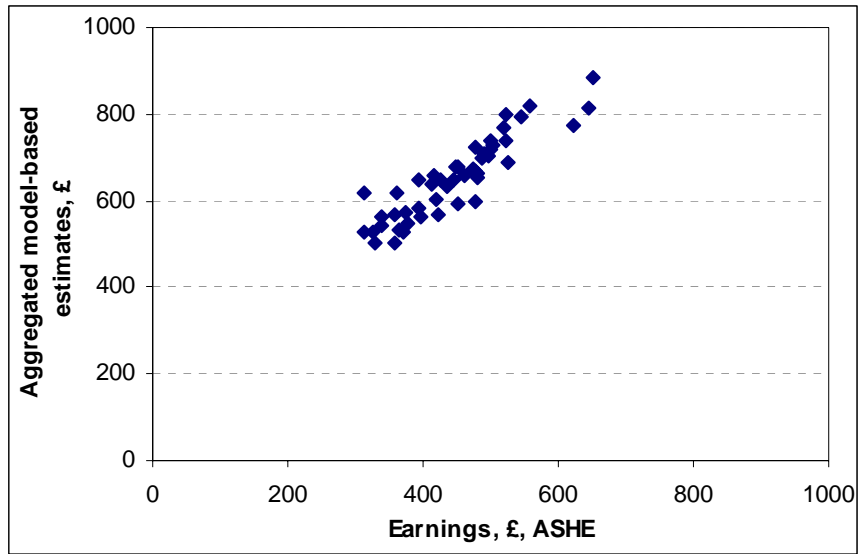


Figure 74: Aggregated model-based estimates compared with ASHE data, East

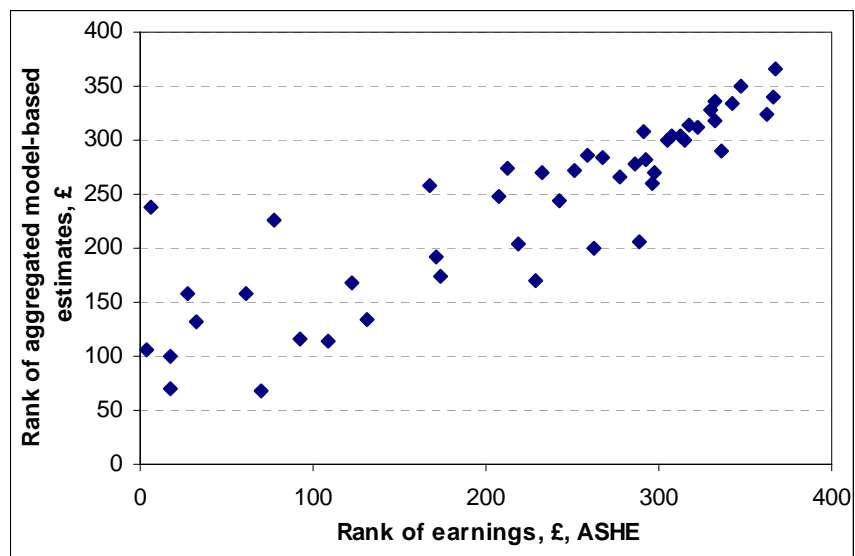


Figure 75: Rank of aggregated model-based estimates compared with rank of ASHE data, East

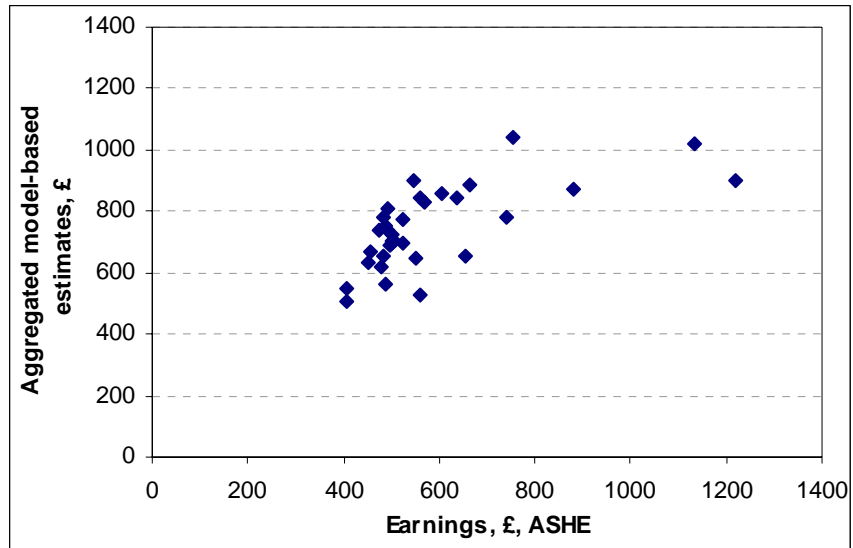


Figure 76: Aggregated model-based estimates compared with ASHE data, London

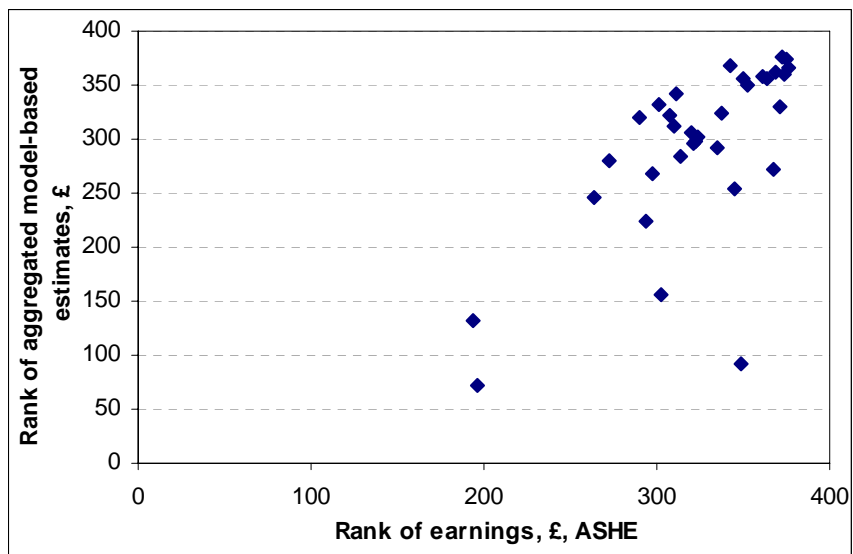


Figure 77: Rank of aggregated model-based estimates compared with rank of ASHE data, London

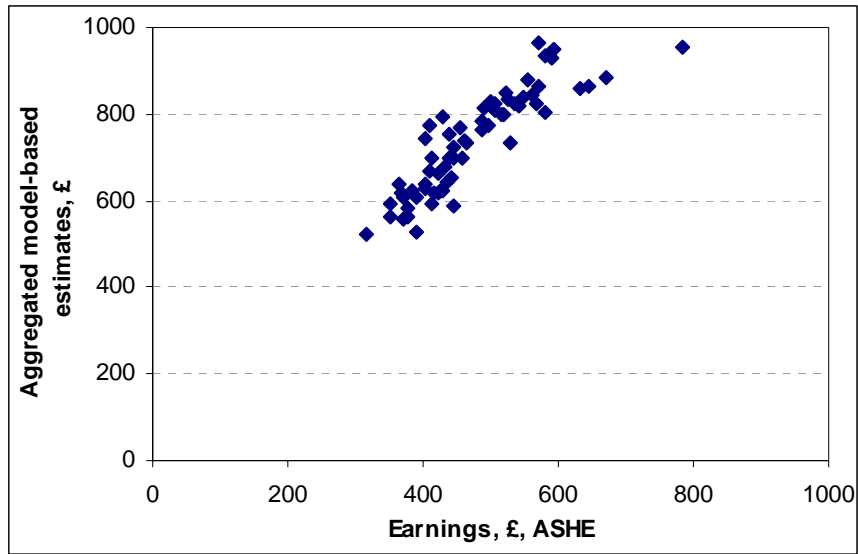


Figure 78: Aggregated model-based estimates compared with ASHE data, South East

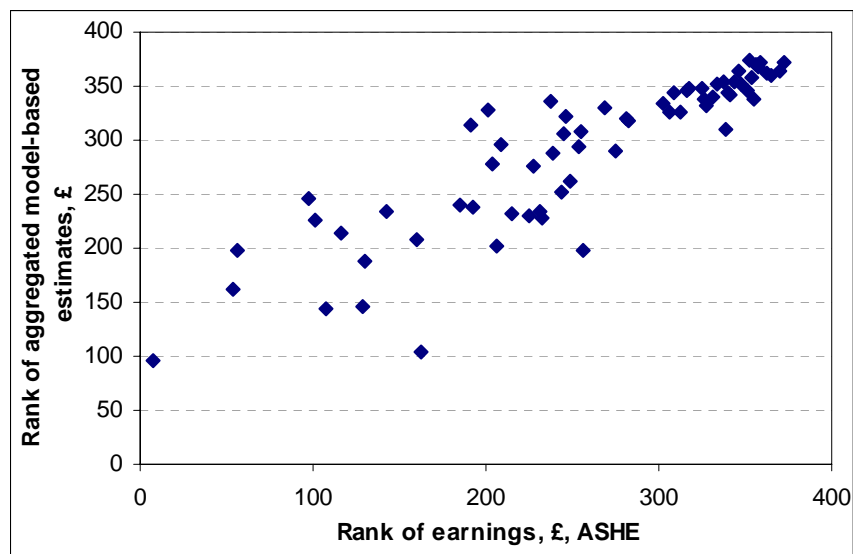


Figure 79: Rank of aggregated model-based estimates compared with rank of ASHE data, South East

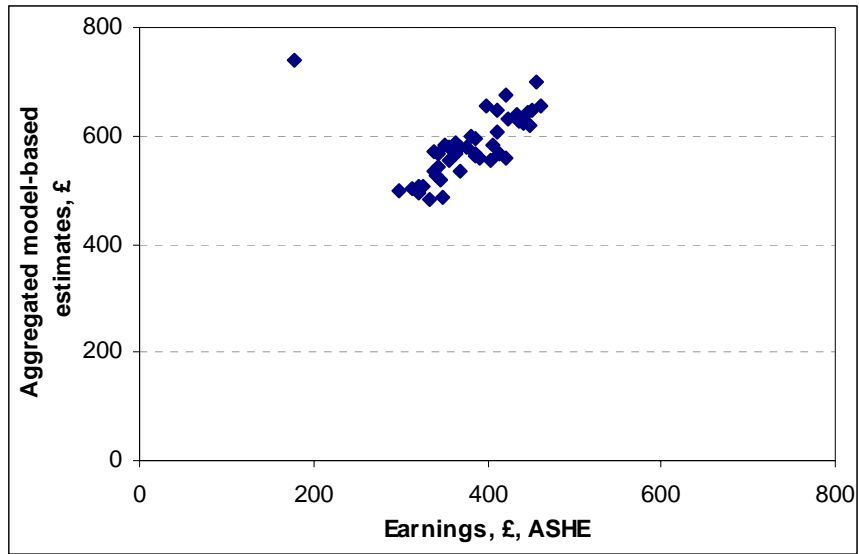


Figure 80: Aggregated model-based estimates compared with ASHE data, South West

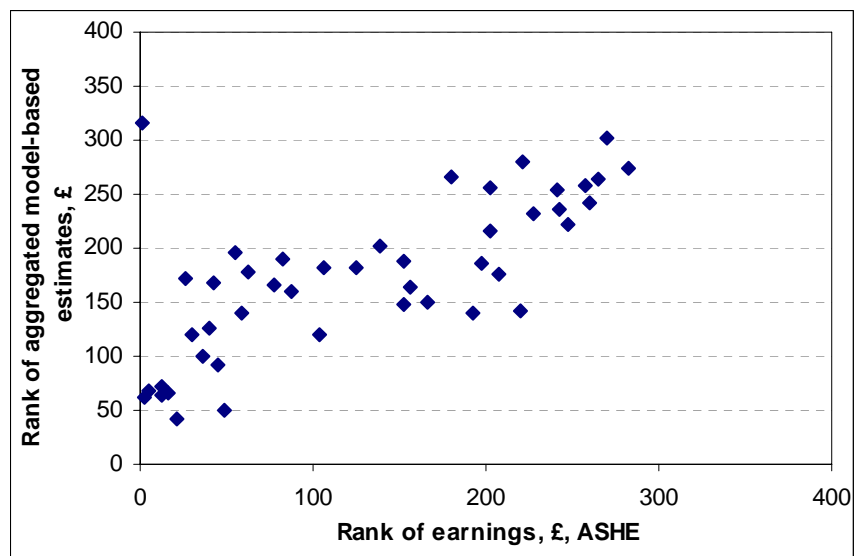


Figure 81: Rank of aggregated model-based estimates compared with rank of ASHE data, South West

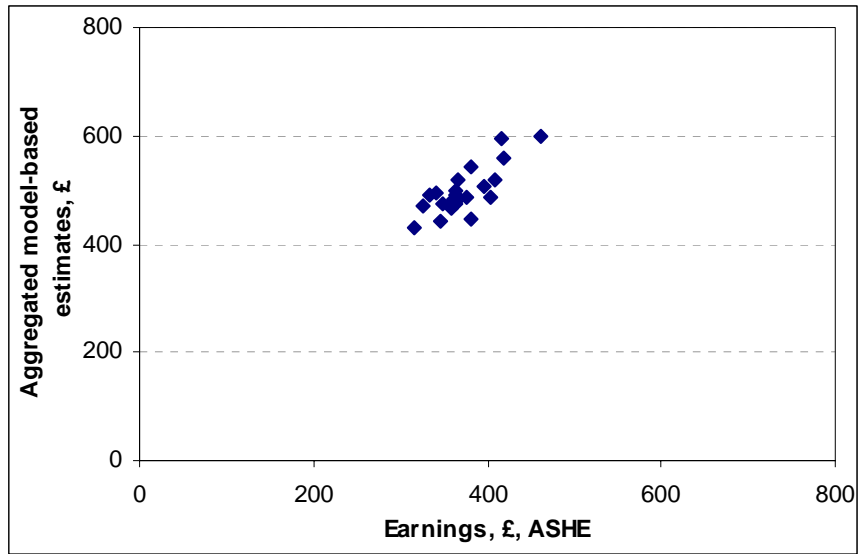


Figure 82: Aggregated model-based estimates compared with ASHE data, Wales

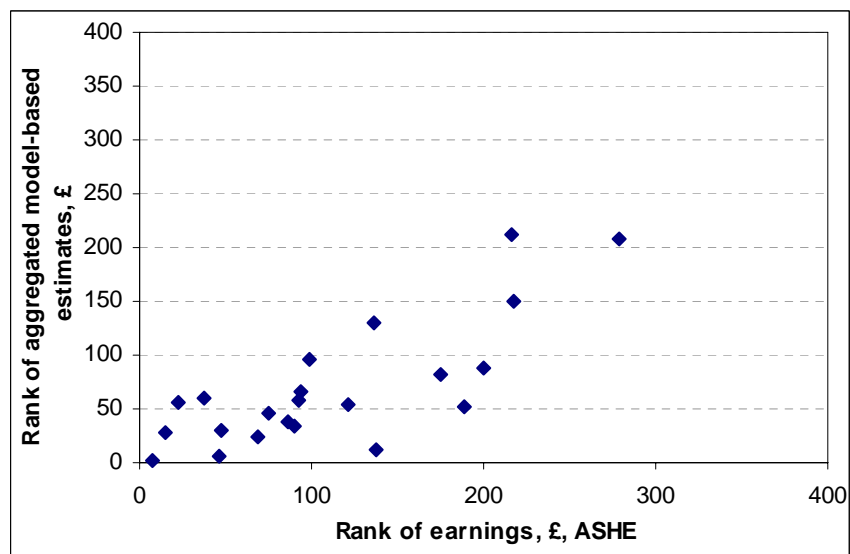


Figure 83: Rank of aggregated model-based estimates compared with rank of ASHE data, Wales

B.5. General Household Survey – England and Wales

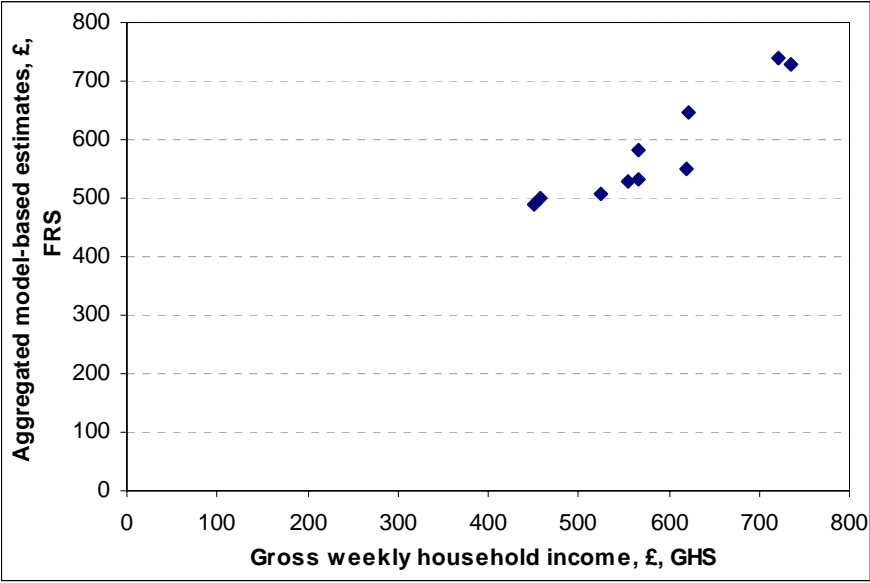


Figure 84: Aggregated model-based estimates compared with GHS data, England and Wales

C Bibliography

Chung, R., Cockerham, J., Ellerd-Elliott, S., Herring, I., Hirani, B., McColl, H., Shome, J., Snow, J. (2006).

Family Resources Survey: Great Britain, 2004-05

http://www.dwp.gov.uk/asd/frs/2004_05/index.asp

Department for Work and Pensions, HBAI team (2006).

Households Below Average Income: 1994/95 – 2004/05.

<http://www.dwp.gov.uk/asd/hbai/hbai2005/contents.asp>

Heady, P., Clarke, P., Brown, G., Ellis, K., Heasman, D., Hennell, S., Longhurst, J., Mitchell, B. (2003).

Small Area Estimation Project Report. Model-Based Small Area Estimation Series No.2, ONS Publication.

Longhurst, J., Cruddas, M., Goldring, S., (2005).

Model-based Estimates of Income for wards 2001/02, Technical Report. Soon to be published in the Model-Based Small Area Estimation Series, an ONS Publication.

Department of Communities and Local Government (2004).

Indices of Deprivation for LSOAs in England, 2004. Source Organisation: Department of Communities and Local Government.

<http://www.communities.gov.uk>

<http://www.neighbourhood.statistics.gov.uk/>

Office for National Statistics (2002).

National Statistics Code of Practice: Protocol on Release Practices.

[http://www.statistics.gov.uk/StatBase/Product.asp?vlnk=9417&Pos=1
&ColRank=1&Rank=256](http://www.statistics.gov.uk/StatBase/Product.asp?vlnk=9417&Pos=1&ColRank=1&Rank=256)

ONS Publication.

Office for National Statistics (2006).

Annual Survey of Hours and Earnings 2005: Labour Market. ONS Publication.

<http://www.statistics.gov.uk/StatBase/Product.asp?vlnk=14203>

Office for National Statistics (2007).

General Household Survey 2005. ONS Publication.

<http://www.statistics.gov.uk/StatBase/Product.asp?vlnk=5756>

Welsh Assembly (2005).

Indices of Deprivation for LSOAs in Wales, 2005. Source

Organisation: Oxford University, produced on behalf of the National Assembly for Wales.

<http://neighbourhood.statistics.gov.uk/>

White, N., Coxon, C., Clarke, P., Cruddas, M., Gask, K., Longhurst, J., Goldring, S. (2007).

Model-based Estimates of Income for MSOA 2004/05, Technical Report. Soon to be published in the Model-Based Small Area Estimation Series, an ONS Publication.